Gene identification and characterization of sex chromosome Y in the liverwort, *Marchantia polymorpha*

Kimitsune Ishizaki

2003

Gene identification and characterization of sex chromosome Y in the liverwort, *Marchantia polymorpha*

Kimitsune Ishizaki

2003

Contents

Abbreviations

CDPK	calcium-dependent protein kinase
dCTP	deoxycytidine 5'-triphosphate
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
FISH	fluorescence *in situ* hybridization
kb	kilobase pairs
Mb	megabase pairs
ORF	open reading frame
PAC	P1-derived artificial chromosome
PCR	polymerase chain reaction
RACE	Rapid amplification of cDNA ends
rDNA	ribosomal DNA
rRNA	ribosomal RNA
RDA	representational difference analysis
RNase	ribonuclease
RT	reverse-transcription
SDS	sodium dodecyl sulfate
SSC	sodium chloride/sodium citrate (buffer)
SSPE	sodium chloride/sodium phosphate/EDTA
STS	sequence-tagged site
TAE	Tris-acetate/EDTA electrophoresis buffer
TE	Tris EDTA
Tris	tris (hydroxymethyl) aminomethane

Introduction

Sex determination and sex chromosome

In evolutionary terms, sexual reproduction has provided genetic variability, and is essential for long-term survival in a complicated and ever-changing environment. Generally, there are only two types of sexual reproduction system. Male and female gametes are produced by the same individual (cosexuality or hermaphrodite) or by separate individuals (unisexuality or dioecy). Most animal species are unisexual (dioecious), whereas the majority of flowering plants are cosexual (hermaphroditic). In dioecious species, there are various sex determination systems. Table 1 provides some of the different sex determination schemes and also show that karyotype determines sex in many dioecious species. For detailed discussion about sex determination systems in animals and plants, see references (Hodgkin, 1992; Marin and Baker, 1998; Ainsworth *et al.*, 1998; Juarez and Banks, 1998).

Table 1. Example of different sex determination systems (reprinted and modified from Hodgkin, 1992)

Species	Mechanism	Sexes	
Mammals	CSD[1] : dominant Y	XX female	XY male
Birds	CSD : ratio?	ZW female	ZZ male
Turtles	ESD[2] : temperature	Warm: female	Cool: male
Alligators	ESD : temperature	Cool: female	Warm: male
Insects			
Drosophila melanogaster	CSD : X/A ratio	XX female	XY male
Musca domestica	CSD : dominant M locus	m/m female	M/m male
Apis mellifera	CSD : haplo-diploidy	Diploid female	Haploid male
Namatodes			
Caenorhabditis elecans	CSD : X/A ratio	XX hermaphrodite	XO male
Meliodogyne incognita	ESD : population density	Sparse: female	Crowded: male
Plants			
Silene latiforia (Angiosperm)	CSD : dominant Y	XX female	XY male
Cannabis sativa (Angiosperm)	CSD : dominant Y	XX female	XY male
Rumex acetosa (Angiosperm)	CSD : X/A ratio	XX female	XY_1Y_2 male
Humulus japonicas (Angiosperm)	CSD : X/A ratio	XX female	XY_1Y_2 male
Salsola komarovii (Angiosperm)	ESD : photoperiod	Shortday: female	Longday: hermaphrodite
Ceratopteris richardii (Fern)	ESD : pheromone[3]	Absent: hermaphrodite	Present: male
Marchantia polymorpha (liverwort)	CSD : dominant Yand/or X ?	X female	Y male

1) CSD, chromosomal sex determination
2) ESD, environmental sex determination
3) The pheromone, antheridiogen, which is secreted into the surroundings by hermaphrodites

Morphologically and genetically distinct X and Y chromosome (with male heterogamety), or Z and W (with female heterogamety), have evolved independently in many groups of animal and plants (Chalesworth, 1996; Charlesworth and Charlesworth, 2000). Sex chromosomes, found in several different groups of organisms, have several common characteristics. Hemizygous chromosomes (that is, the Y chromosome in XY or the W chromosome in ZW) tend to be small, gene poor, except for some genes with functions specific to the heterogametic sex and rich in repetitive sequence. Their non-sex-specific partners, the X chromosome and Z chromosome, tend to be more autosome-like in form and content, and in many cases undergo dosage compensation to equalize amounts of gene products in males and females. It is generally believed that the sex chromosomes have evolved from a pair of ordinary autosomes, and have only gradually diverged (Ellis, 1998; Charlesworth, 1991; Charlesworth, 1996). This gross convergence of sex chromosomes among disparate lineages suggests that similar evolutionary forces have operated in different lineages.

Chromosomal sex determination in animals

In mammals, sex is determined by the presence or absence of the Y chromosome, which encodes the *SRY* gene necessary for testis development (Sinclair *et al.*, 1990; Koopman *et al.*, 1990). In another vertebrate, a candidate sex-determining gene, the Y chromosome-specific *DMY* (DM-related PG17 Y), was identified in the medaka fish, *Oryzias latipes* (Matsuda *et al.*, 2002). The human Y chromosome is small and harbors just a few dozens of genes, while the X chromosome, its meiotic partner, contains several thousand genes. The Y chromosome carries genes which are involved in testis-specific function and have no X-linked homologues (Bachtrog and Charlesworth, 2001). The Y and X chromosomes have evolved independently after recombination between them was suppressed. The X chromosome can still recombine in females, where two X chromosomes can pair, whereas the most of Y chromosome cannot cross over. The lack of recombination over most of the Y chromosome results in the accumulation of deleterious mutations and genetic erosion of the Y chromosome, due to lack of repairing these mutations by homologous recombination. In response, dosage compensation has evolved to maintain equality of the dosage of gene products from X-linked loci in males and females (Charlesworth, 1996; Ellis, 1998; Lahn *et al.*, 2001).

In *Drosophila melanogaster* and *Caenorhabditis elegans,* the ratio of the X chromosomes against the autosomes (the X: autosome dosage) determines their sexes. In *D. melanogaster*, the X chromosome-linked *Sex-lethal* (*Sxl)* gene is at the top of the hierarchy promoting development into female (Marin and Baker, 1998). The Y chromosome of *D. melanogaster* is made up of various types of functionless repetitive sequences, and classified as heterochromatic. Although the Y chromosome of the *D. melanogaster* does not determine its sex, the Y chromosome carries several genes involved in male fertility (Charlesworth, 2001; Carvalho *et al.*, 2000; Carvalho *et al.*, 2001; Reugels *et al.*, 2000; Kurek *et al.*, 1998). In *C. elegans*, the product of the X chromosome-linked *XO lethal-1* (*xol-1*) gene starts a cascade of promoting development into male (Marin and Baker, 1998). Although the evolution of X/autosomal balance sex determination has yet been poorly understood, the sex chromosomes of *D. melanogaster* are also believed to have descended from a pair of ordinary autosomes (Charlesworth, 1996).

Chromosomal sex determination in flowering plants

In plants, the great majority of angiosperms are hermaphrodites, which develop bisexual flowers that carry both male and female reproductive organs, e.g., thale cress (*Arabidopsis thaliana*) and rice (*Oryza sativa*). Sex determination systems in plants have evolved many times from hermaphroditic ancestors, including monoecious plants with separate male and female flowers on the same individual, e.g., maize (*Zea mays*) and cucumber (*Cucumis sativus*), and sex chromosome systems have arisen several times in plant evolution (Charlesworth, 2002). In a dioecious angiosperm, *S. latifolia*, a male individual has an X and a Y chromosomes, while a female individual has two X chromosomes. The *Silene latifolia* Y chromosome dominantly and positively induces male development by its presence in an XX/XY system similar to the mammals. In another dioecious angiosperm, *Rumex acetosa*, a male individual has one X and two Y chromosomes (Y1 and Y2), while a female individual has two X chromosomes. The presence of the Y1 and Y2 chromosomes has no influence on triggering male development, but the ratio of the number of the X chromosome to the number of autosomes (the X: autosome dosage) determines the sex of each individual, as in *D. melanogaster* (Ainsworth *et al.*, 1998; Juarez and Banks, 1998).

Some genes have been isolated from the sex chromosomes of *S. latifolia*. An X-linked gene, *MROS3*, is specifically expressed in male reproductive organs and has a

degenerate homologous sequence on the Y chromosome (Matsunaga *et al.*, 1996; Guttman and Charlesworth *et al.*, 1998; Filatov *et al.*, 2000). Recently it was reported that at least two copies of *MROS3* are present in tandem on the X chromosome (Kejnovsky *et al.*, 2001). Genes *SlY1* and *SlY4* Y are chromosome-linked and transcriptionally active, but there are also active homologues, *SlX1* and *SlX4*, on the X chromosome. These sex chromosome genes are thus not likely to have sex-specific functions such as sex determination (Delichere *et al.*, 1999; Atanassov *et al.*, 2001). Though several studies indicate that sex chromosomes are mostly responsible for sexual reproduction in plants, no sex chromosome gene responsible for sexual reproduction has been identified in any dioecious angiosperms so far.

Chromosomal sex determination in the liverwort, Marchantia polymorpha

In contrast to angiosperms, bryophytes produce not flowers, but archegonia (female sexual organs) and/or antheridia (male sexual organs) on a gametophyte. Commonly bryophytes, especially liverworts are dioecious. A liverwort, *Marchantia polymorpha,* individual forms sexual organs of only one sex and has morphologically distinct sex chromosome (Fig. 1): a male individual has only a Y chromosome (approximately 10 Mb) along with eight autosomes, and a female individual has only an X chromosome (20-25 Mb) along with eight autosomes (Bischler, 1986; Okada *et al.*, 2000). This X/Y exclusiveness strongly suggests the presence of sex determining factors on the sex chromosomes. Furthermore, *M. polymorpha* appears to be the first land plant (Qui *et al.*, 1998), making *M. polymorpha* a unique model to reveal the fundamental mechanism of diverged sexual reproduction systems in land plants.

In terms of evolution, the sexual reproduction system of the liverwort *M. polymorpha*, is very interesting, because they spend most of their life cycles as haploid gametophyte (Fig. 2), while most of higher plants and animals spend most of their life cycles as diploid (polyploid) sporophyte. In higher organism, sexual reproduction can be regarded as a haploid-diploid cycle (HDC) (Tüzel *et al.*, 2001). The life cycle and haploidy of bryophytes is similar to that of green algae and yeast. Research on the sexual reproduction system of bryophyte may uncover a transition point from the primitive land plant to higher plant, fern, which spends most of their life cycle as sporophyte.

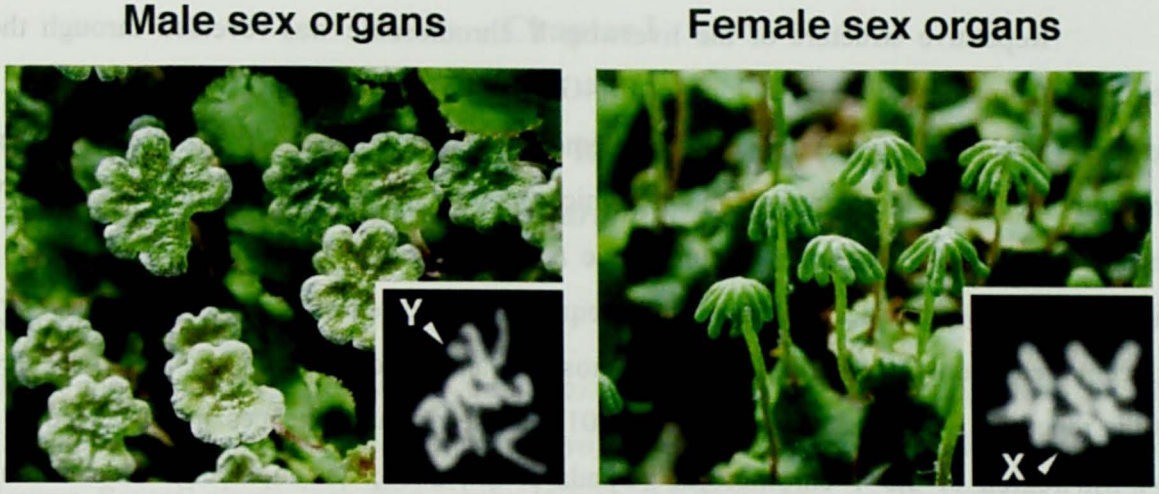


Fig. 1. Chromosomal sex determination of the liverwort, *Marchantia polymorpha*.

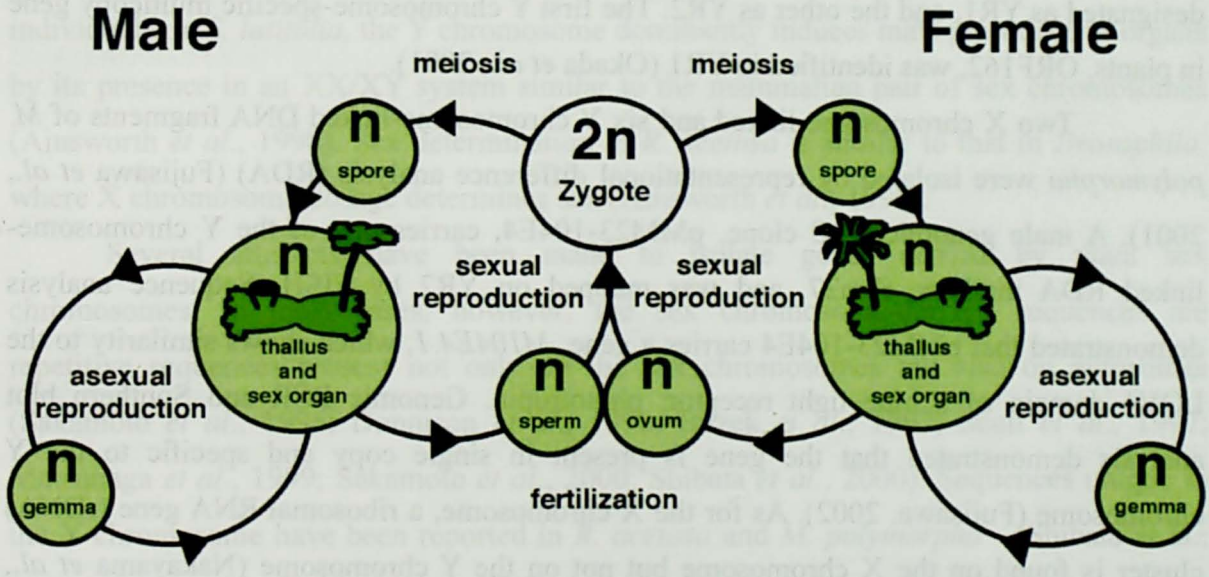


Fig. 2. Life cycle of the liverwort, *Marchantia polymorpha*. *M. polymorpha* spends most of its life cycle as haploid gametophyte (thallus) that can propagate a sexually through gemma. The gametophytes develop sex organs for sexual reproduction in male and female individuals, respectively. *M. polymorpha* is diploid during the limited period of the time right after fertilization. After meiosis, spores are formed on the female sex organ and the spores develop to gametophyte.

Structure of the sex chromosomes in M. polymorpha

Towards a better understanding of the sexual reproduction system in *M. polymorpha*, we have initiated detailed analyses on the gene structure of the Y chromosome. Genomic libraries for male and female *M. polymorpha* plants have enabled us to identify sequences unique to the Y chromosome (Okada *et al.*, 2000).

Repetitive structure of the liverwort Y chromosome was revealed through the analysis of a male-specific PAC clone, pMM4G7, and pMM23-130F12 (Okada *et al.*, 2001). Several Y chromosome-specific sequence elements of approximately 70 to 400 nucleotides are combined into larger arrangements, which in turn are assembled into extensive Y chromosome-specific stretches. Fluorescence *in situ* hybridization (FISH) analysis using the Y chromosome-specific repetitive sequence indicated that the sequences are concentrated in only one half of the Y chromosome, and the other half remains largely free of the repetitive sequences (Okada *et al.*, 2001). These repeat sequences are dispersed in a 3-4 Mb region of the Y chromosome (Okada *et al.*, 2001; Okada, 2002). This finding revealed that the Y chromosome can be roughly divided into two distinct segments: A region in which the Y chromosome-specific repetitive sequences are accumulated, is designated as YR1, and the other as YR2. The first Y chromosome-specific multicopy gene in plants, ORF162, was identified in YR1 (Okada *et al.*, 2001).

Two X chromosome-linked and six Y chromosome-linked DNA fragments of *M. polymorpha* were isolated by representational difference analysis (RDA) (Fujisawa *et al.*, 2001). A male genomic PAC clone, pMM23-104E4, carries one of the Y chromosome-linked RDA markers, rbm27, and was mapped on YR2 by FISH. Sequence analysis demonstrated that pMM23-104E4 carries a gene, *M104E4.1*, which shows similarity to the LOV1 domain of a blue light receptor, phototropin. Genomic PCR and Southern blot analysis demonstrated that the gene is present in single copy and specific to the Y chromosome (Fujisawa, 2002). As for the X chromosome, a ribosomal RNA gene (rDNA) cluster is found on the X chromosome but not on the Y chromosome (Nakayama *et al.*, 2001).

In this thesis, the author has intended to improve our knowledge on genes and evolution of the liverwort, *M. polymorpha*. In Chapter I, sequence analysis of a part of YR1, identified several genes and their unique feature. In Chapter II, sequence analysis of a 0.7-Mb region of YR2 suggested difference in gene organization between YR1 and YR2. In Chapter III, a diverged homologue of a Y chromosome-linked gene embedded on the X chromosome was characterized through sequence analysis of the X chromosome-derived PAC clone.

Chapter I

Multicopy genes uniquely amplified in the Y chromosome-specific repeats of the liverwort, *Marchantia polymorpha*

INTRODUCTION

Unlike most animal species, which are unisexual, the majority of flowering plants such as *Arabidopsis thaliana* and rice (*Oryza sativa*) are hermaphroditic and develop bisexual flowers. Several plants, including gingko (*Gingko biloba*), white campion (*Silene latifolia*), garden sorrel (*Rumex acetosa*), hemp (*Cannabis sativa*) and the liverwort, *Marchantia polymorpha,* are dioecious, that is, unisexual reproductive organs are formed on different individuals. In *S. latifolia*, the Y chromosome dominantly induces male reproductive organs by its presence in an XX/XY system similar to the mammalian pair of sex chromosomes (Ainsworth *et al.*, 1998). Sex determination in *R. acetosa* is similar to that in *Drosophila,* where X chromosome dosage determines sex (Ainsworth *et al.*, 1998).

Several attempts have been made to isolate genes carried by plant sex chromosomes. In most cases, however, the sex chromosome-derived sequences are repetitive sequences present not only on the sex chromosomes but also on autosomes (Sakamoto *et al.*, 1995; Donnison *et al.*, 1996; Buzek *et al.*, 1997; Scutt *et al.*, 1997; Matsunaga *et al.*, 1999; Sakamoto *et al.*, 2000; Shibata *et al.*, 2000). Sequences unique to the Y chromosome have been reported in *R. acetosa* and *M. polymorpha* (Shibata *et al.*, 1999; Okada *et al.*, 2001). In terms of active genes carried by plant sex chromosomes, the *S. latifolia MROS3* gene was originally identified as a gene specifically expressed in male reproductive organs (Matsunaga *et al.*, 1996). This gene was later shown to be X chromosome-linked with a homologous sequence present on the Y chromosome (Guttman *et al.*, 1998). Recently it was reported that at least two copies of *MROS3* are present in tandem on the X chromosome (Kejnovsky *et al.*, 2001). Genes *SlY1* and *SlY4* are found to be Y chromosome-linked and active in *S. latifolia,* but there are also close active homologues, *SlX1* and *SlX4*, on the X chromosome. These sex chromosome genes are therefore not likely to have sex-specific functions such as sex determination (Delichere *et*

al., 1999; Atanassov *et al.*, 2001). Genes responsible for sex determination in plants, as well as the detailed structures of the plant sex chromosomes are still largely unresolved.

The liverwort, *M. polymorpha*, has unusually small sex chromosomes. Because of its haploidy Y chromosome is present only in male plants and morphologically distinct X chromosome is found only in female plants (Okada *et al.*, 2000). This X/Y exclusiveness strongly suggests the presence of sex determining factors on the sex chromosomes. Towards a better understanding of the sexual reproduction system in *M. polymorpha*, we have initiated detailed analyses on the gene structure of the Y chromosome. Genomic libraries for male and female *M. polymorpha* plants have enabled us to identify sequences unique to the Y chromosome (Okada *et al.*, 2000). Here we describe five multicopy genes amplified in a highly repeated region of the Y chromosome.

MATERIALS AND METHODS

Plant materials

Male and female thalli of *M. polymorpha* (E lines) (Takenaka *et al.*, 2000) were cultivated on M51C medium (Ono *et al.*, 1979) at 24°C under continuous light. Sex organs were obtained from the same lines.

Sequence analyses

Shotgun sequencing of pMM2D3 was performed as described previously (Okada, *et al.*, 2001). Searches for protein coding regions were performed against the non-redundant protein sequence database at the National Center for Biotechnology Information (NCBI) using the BLASTX program (Altschul *et al.*, 1990) and against *M. polymorpha* ESTs (Nagai *et al.*, 1999; Nishiyama *et al.*, 2000) using BLASTN program (Altschul *et al.*, 1990).

The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF542555, AF542560, AF542556, AF542557, AF542558, and AF542559 for RS, RT, R1, R2, R3, and R4 portions in Figure 1, respectively.)

Genomic PCR analysis

PCR for determining male-specificity of putative genes was performed basically as described previously (Okada *et al.*, 2000). Template genomic DNAs were isolated as described by Takenaka *et al.* (Takenaka *et al.*, 2000), and 10 ng each were used as template. Sequences of primers are, 5'-CTGGACCAAGTGATTCGCTCTC and
5'-AGCCCACTGATATAACGAAGAC for M2D3.1,
5'- CCGTGACGCCGAGCGATGTGGG and
5'- CGCTCGAACGACACCGTATCGC for M2D3.2,
5'-GGAATGCATCCCAGTTGAGACCand
5'- AAGAGCCTCGAGCTTCTGCTTC for M2D3.3,
5'-CGGACTGGAGTACTGGAACGAT and
5'-TTCTGGTCGGAACTGCTGATCG for M2D3.4, and
5'-AAACTTTCGCTGCATCGAGCGG and
5'- TCGTCCTGTTTCTGCTTCAGCC for M2D3.6. Primers designed from ORF162 (Okada *et al.*, 2001) and the calcium-dependent protein kinase (CDPK) gene (Okada *et al.*, 2000) were used for quality evaluation of the genomic DNA.

Southern blot analysis

Five micrograms of genomic DNA were digested with *Bam*HI and the resulting fragments were separated in a 1% agarose gel in 1 x TAE buffer. After alkaline treatment and blotting onto a nylon membrane, hybridization was performed in a solution containing 5 x Denhardt's reagent, 6 x SSPE pH 7.4, 0.5% SDS, 100 μg/ml denatured salmon sperm DNA, and 50% formamide at 42°C. To prepare probes, DNA fragment specific to each gene was amplified by PCR with plasmid DNA of pMM2D3 as template and with the primer pairs described in genomic PCR analysis (Materials and Methods), and then the PCR products were labeled with [α-^{32}P] dCTP by another round of PCR. Membranes were washed for 1 hr in a solution containing 1 x SSPE and 0.1% SDS at 42°C followed by two washes with 0.1 x SSPE and 0.1% SDS at 65°C for 1 hr. Radioactive signals were visualized with a BAS2000 Image Analyzer (Fuji Photo Film).

RT-PCR analysis

Total RNAs from male and female thalli, and from male and female sex organs were individually prepared by the phenol/SDS method (Ausubel *et al.*, 1987). Poly(A)+ RNA was prepared with the PolyATract™ System 1000 (Promega) according to the manufacturer's instructions. First-strand cDNA was synthesized from 5 μg of DNase-treated poly(A)+ RNA using SuperScript II™ reverse transcriptase (Gibco BRL) at 42 °C with XhoSseEcoR-dT primer (5'-GAGAATTCCTGCAGGCTCGAGTTTTTTTTTTTTTTTT-3') for 60 min. A 20 μl reaction mixture was diluted in 400 μl TE, and 1 μl of the diluted mixture was used as template in a 20 μl PCR amplification mix containing 10 pmol of the same primers used for the genomic PCR. Reactions without reverse transcriptase were performed to check genomic DNA contaminations.

Northern blot analysis

Poly(A)+ RNA was prepared according to the method described above for RT-PCR. Five micrograms of poly(A)+ RNA were electrophoresed in a 0.8% denaturing agarose gel containing formaldehyde and transferred onto a nylon membrane. Hybridization was performed with ExpressHyb™ solution (CLONTECH) as described (Okada *et al.*, 2001). Part of M2D3.4 was amplified by PCR with plasmid DNA of pMM2D3 as template and the M2D3.4 primer pair. The PCR product of M2D3.4 was labeled with [α-^{32}P] dCTP by another round of PCR. The hybridized membrane was washed for 1 hr in a solution containing 2 x SSC and 0.1% SDS at 25°C followed by two washes with 0.1 x SSC and 0.5% SDS at 55°C for 1 hr. Radioactive signals were visualized with a BAS2000 Image Analyzer (Fuji Photo Film).

RESULTS

A male-specific clone from the Y chromosome

To identify Y chromosome-derived clones in our PAC (P1-derived artificial chromosome) library, we screened for clones containing the Y chromosome-specific repeat sequences (Okada *et al.*, 2000). One positive clone, pMM2D3, was selected for detailed analysis,

since the restriction profile of pMM2D3 is distinct from that of pMM4G7, which has been already examined (Okada *et al.*, 2001), and there are other PAC clones that align consistently with pMM2D3 (data not shown), excluding a possibility that pMM2D3 is a chimeric clone.

Approximately 1,800 shotgun and gap-filling sequences yielded an over 8 times coverage of the entire length of the clone (approximately 90 kb long) and assembly resulted in six sequence contigs. The sequences of the two ends of the insert, RS and RT, were unambiguously oriented by the *Sfi*I and *Pac*I recognition sites (Fig. 1-1). The total length of R1, R2, R3, and R4 (approximately 18 kb) is shorter than the actual distance between contigs RS and RT (approximately 26 kb) because the copy numbers of the repeat sequences are unknown. Nevertheless, any of the 1,800 shotgun and gap-filling sequences aligned with one of the six contigs, indicating that the entire sequence content of pMM2D3 is represented in the six contigs.

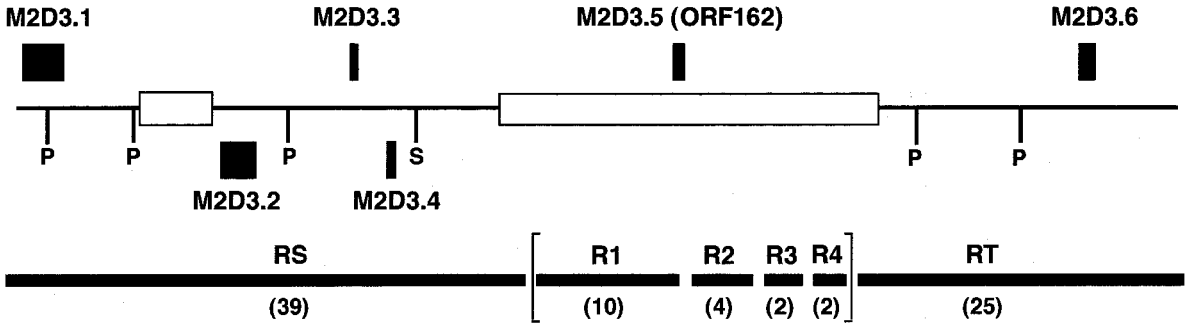


Fig. 1-1. Schematic diagram of the structure of pMM2D3. The insert of the plasmid is displayed with its SP6-end on the left. Alignments of sequence contigs are shown by the lower horizontal bars labeled as RS, RT, R1, R2, R3, and R4. Sequences of R1, R2, R3, and R4 consisted of the Y chromosome-specific repeat sequences. Approximate sizes of contigs are given in parentheses in kb. The order of contigs in brackets was not determined. Putative genes are indicated by closed boxes above (left-to-right orientation) and below (right-to-left orientation) the line. Recognition sites for *Pac*I (P) and *Sfi*I (S) determined by restriction digestion and electrophoresis are indicated by vertical lines. Clusters of the Y chromosome-specific repeat sequences are indicated by open boxes. The exact position of M2D3.5 within the clone is not known because of the repeat sequences, but its orientation is predicted from the orientation of the repeat sequences at the proximal ends of RS and RT.

Y chromosome-specific repeat sequences in pMM2D3

The Y chromosome-specific repeat sequences identified in clone pMM4G7 are also highly conserved in pMM2D3. In addition to the five types of Y chromosome-specific *Bam*HI fragments (1.5, 2.4, 2.7, 5.2, and 2.2 kb) identified in clones pMM4G7 and pMM23-130F12 (Okada *et al.*, 2001), two *Bam*HI variants with different sizes, 2.0 and 1.8 kb, were found in pMM2D3 (Fig. 1-2). The 2.0-kb *Bam*HI fragment has one *Mbo*I element while the 2.4-kb *Bam*HI fragment has two, and the 1.8-kb *Bam*HI fragment has three additional copies of the *Hae*III element compared to the 1.5-kb *Bam*HI fragment (Fig. 1-2). The 2.0-kb and 1.8-kb *Bam*HI fragments found in pMM2D3 thus also consist of common sequence elements and differ only in relative copy numbers of these elements.

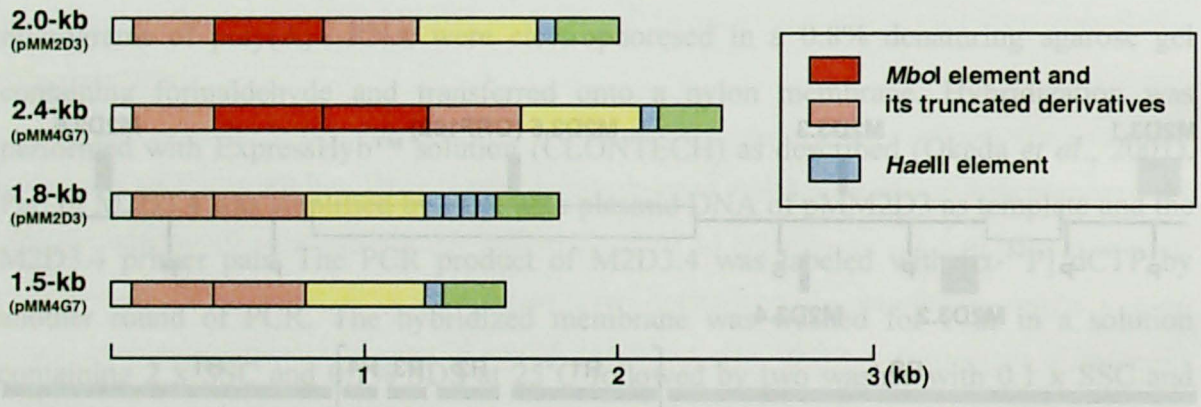


Fig. 1-2. Structures of the 2.0-kb and 1.8-kb *Bam*HI repeat units newly identified in pMM2D3, in comparison to the 2.4-kb and 1.5-kb *Bam*HI repeat units of pMM4G7. As in the 2.4-kb and 1.5-kb *Bam*HI repeat units, the 2.0-kb and 1.8-kb *Bam*HI repeat units also consist of common subrepeats, *Mbo*I (red) and *Hae*III elements (blue). Other sequences are color-coded (white, yellow, and green) according to their respective similarities.

Potential genes found in pMM2D3

In order to identify protein-coding genes, the sequence of pMM2D3 was searched against the 1,415 *M. polymorpha* ESTs (Nagai *et al.*, 1999; Nishiyama *et al.*, 2000) and the non-redundant protein sequence database of NCBI. No *M. polymorpha* EST tags any portion of pMM2D3. However, five regions, M2D3.1, M2D3.2, M2D3.3, M2D3.4 and M2D3.6 of pMM2D3, show significant similarity of their deduced amino acid sequences to known proteins (Fig. 1-1 and Table 1-1). These are in addition to M2D3.5, which is a member of the previously described Y chromosome-specific ORF162 gene family (Okada *et al.*, 2001). All these regions with the exception of M2D3.1 align as uninterrupted open reading frames (ORFs) and thus appear to be intact protein-coding genes.

Table 1-1 Potential genes found in pMM2D3

Contig	Sequence	Length (bp)	% Identity (aa)	Similar sequences (species)	Accession No.
RS	M2D3.1	1173[a)]	67% (65/96)	putative alliinase (*Arabidopsis thaliana*)	T05567
RS	M2D3.2	2073	68% (76/111)	putative RAV-like B3 DNA binding protein (*Arabidopsis thaliana*)	AAC34233
RS	M2D3.3	492	35% (65/184)	unknown protein (*Arabidopsis thaliana*)	AAG51648
RS	M2D3.4	588	33% (35/105)	LGC1 (*Lilium logiflorum*)	AAD19962
R1	M2D3.5	489	100% (162/162)	ORF162 (*Marchantia polymorpha*)	BAB62538
RT	M2D3.6	367	47% (158/334)	unknown protein (*Oryza sativa*)	AAG03087

a) Putative introns were not included.

The segmented alignment of amino acid sequences of M2D3.1 and alliin lyase gene homologues indicates that M2D3.1 consists of six exons. The joined sequence of the six putative exons contains an uninterrupted ORF. The dinucleotides GT and AG characteristic for 5'- and 3'-ends of eukaryotic introns are conserved at the boundaries of the deduced introns. The alliin lyase gene of onion has four introns, all of which are located at positions identical to the putative introns in M2D3.1 (data not shown). In Chinese chive (*Allium tuberosum*) homologue, Lys^{280} has been shown to be essential for catalytic activity and to be a probable pyridoxal-phosphate-binding residue (Manabe *et al.*, 1998), the corresponding lysine residue is conserved in M2D3.1 at position 224 (Fig. 1-3). This suggests that the protein encoded in M2D3.1 could function as an *S*-alk(en)yl-L-cysteine sulfoxide lyase in the metabolism of cysteine, homocysteine and methionine or derivative compounds. Alliin lyase is thought to contribute to the chemical defense in *Allium* plants by producing volatile sulfur-containing compounds (Manabe *et al.*, 1998). The homologous gene in *M. polymorpha* may have an analogous function.

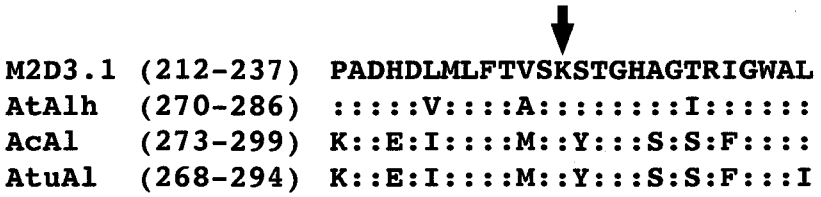
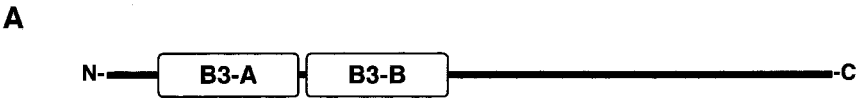


Fig. 1-3. Multiple amino acid sequence alignment of the pyridoxal phosphate binding site in M2D3.1 and its related alliin lyase-like proteins. The numbers in parentheses indicate the positions in the respective sequence. Amino acid residues identical to those of M2D3.1 are indicated by colons. The lysine residue suggested to be the pyridoxal phosphate binding site is indicated by an arrow. AtAlh, alliin lyase homologue of *Arabidopsis thaliana* (GenBank accession number T05567); AcAl, *Allium cepa* alliin lyase (AAA32639); AtuAl, *Allium tuberosum* alliin lyase (BAA20358).

The protein encoded by M2D3.2 contains two B3 DNA-binding domains in tandem orientation. B3 DNA-binding domains are found in a variety of transcription factors such as auxin response factors (ARFs), the maize transcription factor VIVIPAROUS1 (VP1), the RAV1 DNA-binding protein of *A. thaliana* and their relatives (Ulmasov *et al.*, 1997; Suzuki *et al.*, 1997; Kagaya *et al.*, 1999). The B3 DNA-binding domains of M2D3.2 show the highest similarity to that of RAV1 (Kagaya *et al.*, 1999, Fig. 1-4). The B3 DNA-binding domain of RAV1 recognizes the DNA motif CACCTG (Kagaya *et al.*, 1999), but its target genes have not been identified. M2D3.2 is unique in containing two B3 DNA-binding domains (Fig. 1-4), which are 95 % identical in their amino acid sequences.



B

```
B3-A (65-155)  TMQRE--HLFEKAVTPSDVGKLNRLVIPKQHAERCF-PLDLALNVP--CQTLSFEDVSGKHWRFRYSYWNSSQSYV-FTKSWSCFLKGKKLEAGDTV
B3-B (192-282) ARN::--:::::::::::::::::::::::::::::::-:::::::A:--:::::::::::::::::::::::::::-:::G:::::::::::::::
pB3  (28-118)  VAD::--:M:D:V:::::::::::::::::::::::F:-:::SSS:EK--GLL:N:::LT::S:::::::::::::::-M::G::R:V:D:::D:::I:
RAV1 (181-272) FRSA:--A:::::::::::::::::::::::H:::KH:-::PS-S::SVKGVL:N::::N::V:::::::::::::::-L::G::R:V:E:N:R:::V:
RAV2 (181-272) LKT::--V:::::::::::::::::::::::KH:-::PSPSPAVTKGVLIN::::N::V:::::::::::::::-L::G::R:V:E:N:R:::V:
ARF9 (107-200) EL::PKV:S:S:VL:A::TSTHGGFSVLRK::TE:LP:::MTQQT:--T:E:VA:::H:YQ:K:KHIFRGQPRRHL-L:TG::T:VTS:R:V::::F
VP1  (510-600) DKNLR--F:LQ:V:KQ::::S:G:I::::KE::VH:-:ELKTRDGI--SIPME-DIGTSRV:NM::RF:PNNK:RMYLLENTGE:VRSNE:QE::FI
```

Fig. 1-4. Alignment of B3 DNA binding domains found in M2D3.2. **(A)** Schematic illustration of M2D3.2 with two B3 DNA-binding domains, B3-A and B3-B. **(B)** Amino acid sequence alignment of B3 DNA-binding domains found in M2D3.2 and other proteins. The numbers in parentheses indicate the positions in the respective sequences. Amino acid residues identical to those of B3-A, are indicated by colons. Gaps are indicated by dashes. pB3, putative DNA binding protein of *A. thaliana* (GenBank accession number AAC34233); RAV1, related to ABI3/ VP1 1 DNA-binding protein of *A. thaliana* (BAA34250); RAV2, related to ABI3/ VP1 2 DNA-binding protein of *A. thaliana* (BAA34251); ARF9, auxin response factor 9 of *A. thaliana* (AAD24427); VP1, VIVIPAROUS1 protein of maize (P26307).

Multiple alignment of the predicted product of M2D3.6 and the other proteins which show significant similarity to it, revealed that they contain a PAH (paired amphipathic helix repeat) domain which is important for Sin3 function as a corepressor (Fig. 1-5). Similarity of M2D3.6 is limited to the PAH regions in these homologues. While Sin3 and mSin3A have four PAH domains (Brubaker *et al.*, 2000), M2D3.6 has one PAH domain. As PAH domains have been reported to associate with various transcription factors (Brubaker *et al.*, 2000), M2D3.6 potentially interacts with a transcription factor and regulates its activity.

```
M2D3.6   (91-165)   SSASDQTINFINKVKTRFSADEHVYKAFLEILNMYRKGNK------------PISEMYQEVATLFSEHADGEHADLLEEFTSFRPDS
OsUP     (103-171)  PVDFMEA:::V::I:A::QQED::::S::G:::::LH::-------------S:QDV:G::P-::RDYP:-----:::::KH:L::T
AtTRl    (131-200)  TVEFEEA:S:V::I:T::QHN:L:::S:::::::::D::-------------D:T:V:N::ST::ED:S:-----:::::R:L::S
MmSin3A  (301-383)  PVEFNHA::YV::I:N::QGQPDI::::::::HT:Q:EQRNAKEAGGNYTPALTEQ:V:AQ::R::KNQE:-----::S::GQ:L::A
ScSin3   (405-474)  DVEFS:::SYV::I::::ADQPDI::H:::::QT:QREQ:-------------::N:V:AQ:TH::QNAP:-----:::D:KK:L::S
```

Fig. 1-5. Amino acid sequence alignment of the PAH domains found in M2D3.6 and other proteins. The numbers in parentheses indicate the positions in the respective sequence. Amino acid residues identical to those of PAH domain of M2D3.6 are indicated by colons. Gaps are indicated by dashes. OsUP, unknown protein of rice (GenBank accession number AAG03087); AtTRl, transcription regulator-like protein of *Arabidopsis thaliana* (T51447); MmSin3A, Sin3 transcription regulator homologous protein of *Mus musculus* (AAA89119); ScSin3, Sin3 transcription regulator protein of *Saccharomyces cerevisiae* (RGBYS3).

Analysis of sex specificity of the five putative genes found in pMM2D3

To investigate their sex-specificities, we performed diagnostic genomic PCR of the five putative genes (Fig. 1-6A). A primer pair was designed for the conserved region of each of the putative genes, M2D3.1, M2D3.2, M2D3.3, M2D3.4 and M2D3.6. M2D3.4 did not yield a PCR product for female DNA. Three additional different primer sets for M2D3.4 again yielded no PCR products (data not shown). Primer sets of the other four genes readily detected homologous sequences in the female DNA. This result indicates that at least one homologous sequence for each of the putative genes, M2D3.1, M2D3.2, M2D3.3 and M2D3.6, is present on the X chromosome and/or autosomes.

The five putative genes are present in multicopy on the Y chromosome

In order to investigate the copy number of the five putative genes, genomic Southern blot analysis was performed. The five putative genes showed at least one intense signal in the male but not the female DNA, revealing that the male genome carries numerous copies of these putative genes on the Y chromosome (Fig. 1-6B, indicated by closed circles). In contrast to M2D3.5 (ORF162), which as expected shows no signal in the female DNA (Okada *et al.*, 2001), the probes for the five putative genes gave several weakly hybridizing fragments in the female DNA (Fig. 1-6B, indicated by open circles), indicating the existence of homologous sequences on the X chromosome and/or autosomes.

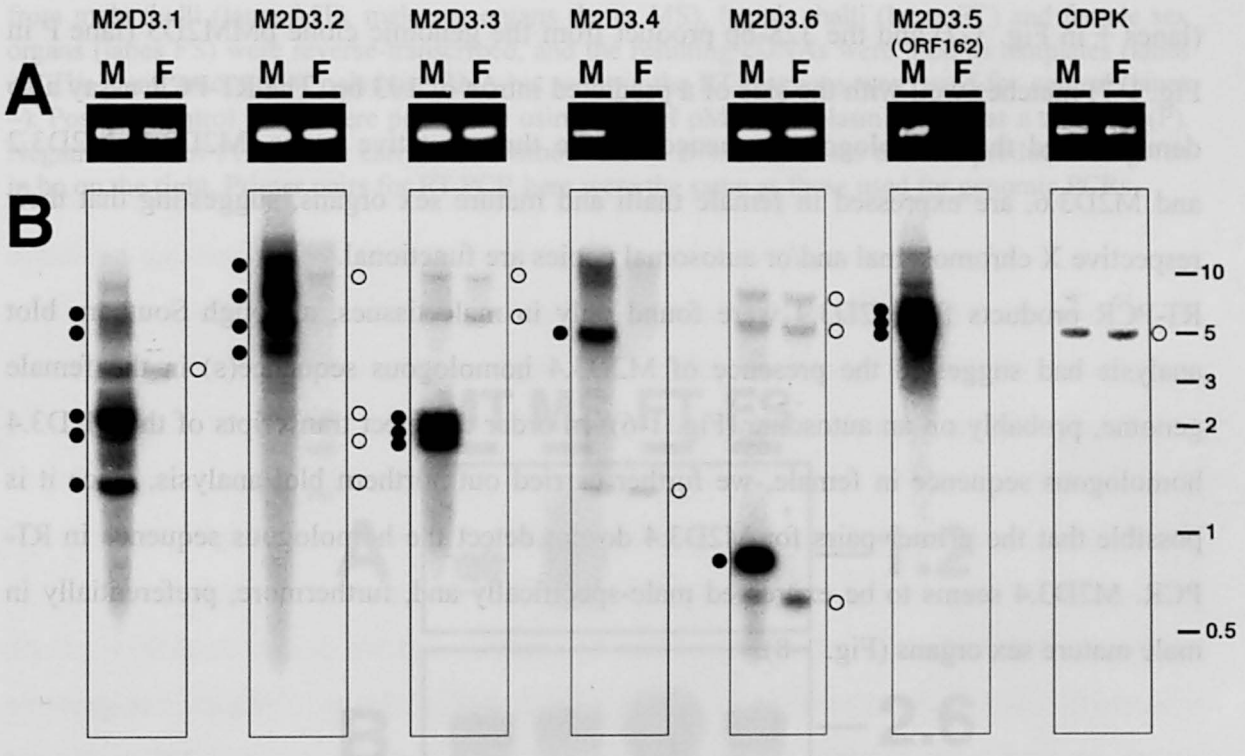


Fig. 1-6. The sex specificity and copy number of the five putative genes. **(A)** Sex specificity of the five putative genes were examined by genomic PCR with primer pairs designed for the conserved regions of the respective five genes. CDPK served as a quality control for the genomic DNAs. **(B)** The copy numbers of the five putative genes were examined by genomic Southern blot analyses. Genomic DNAs of male (M) and female (F) plants were digested with *Bam*HI, and were probed with DNA fragments of the respective ORFs and CDPK. Closed circles indicate intense signals detected specifically in male DNAs. Open circles indicate signals detected in both male and female DNAs. Sizes of signals are given in kb on the right.

Since all the DNA fragments detected in the female DNA are also found in the male DNA (Fig. 1-6B, indicated by open circles), these homologous sequences are likely to be located on autosomes. *A. thaliana* genes similar to M2D3.2 and M2D3.6 have a few diverged homologues on different *A. thaliana* chromosomes but do not appear to be multicopy genes like the liverwort genes. *A. thaliana* genes similar to the other liverwort genes, M2D3.1, M2D3.3, M2D3.4, and M2D3.5, are single-copy genes.

Transcription analysis of the novel putative genes

Expression analyses of the five putative genes were carried out by diagnostic RT-PCR. Except for M2D3.3, four of the putative genes, M2D3.1, M2D3.2, M2D3.4 and M2D3.6, were detected to be expressed in male thalli as well as in mature male sex organs (Fig. 1-7). The size discrepancy of 103 bp for M2D3.1 between the 225-bp PCR products from cDNA (lanes + in Fig. 1-7) and the 328-bp product from the genomic clone pMM2D3 (lane P in Fig. 1-7) matches well with the size of a predicted intron of 103 bp. The RT-PCR assay also demonstrated that homologous sequences of the three putative genes, M2D3.1, M2D3.2 and M2D3.6, are expressed in female thalli and mature sex organs, suggesting that their respective X chromosomal and/or autosomal copies are functional.

RT-PCR products for M2D3.4 were found only in male tissues, although Southern blot analysis had suggested the presence of M2D3.4 homologous sequence(s) in the female genome, probably on an autosome (Fig. 1-6). In order to detect transcripts of the M2D3.4 homologous sequence in female, we further carried out northern blot analysis, since it is possible that the primer pairs for M2D3.4 do not detect the homologous sequence in RT-PCR. M2D3.4 seems to be expressed male-specifically and, furthermore, preferentially in male mature sex organs (Fig. 1-8).

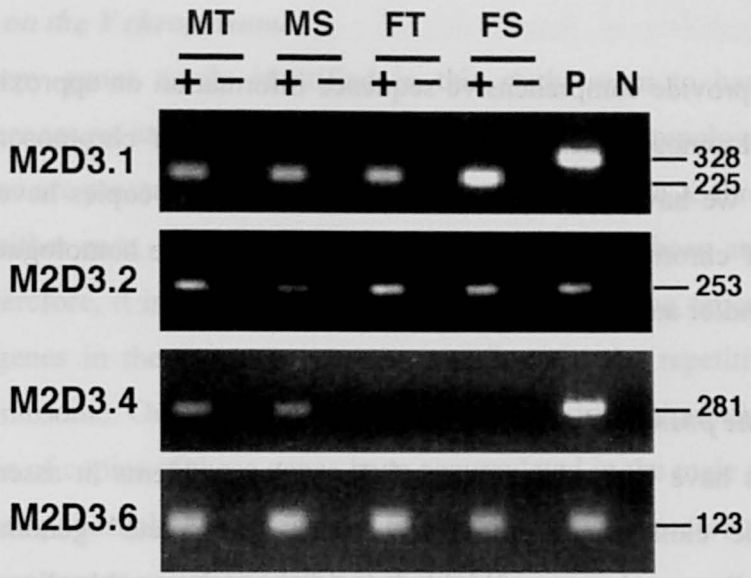


Fig. 1-7. Reverse transcription (RT)-PCR analyses of the putative genes. Poly(A)+ RNAs isolated from male thalli (lanes MT), male sex organs (lanes MS), female thalli (lanes FT) and female sex organs (lanes FS) were reverse-transcribed, and the resulting cDNAs were used as templates (lanes +). The same amounts of poly(A)+ RNA but without the RT reactions were used for control (lanes –). Positive control PCRs were performed using 1 ng of pMM2D3 plasmid DNA as a template (P). Negative control PCRs were carried out without added DNA (N). Sizes of PCR products are given in bp on the right. Primer pairs for RT-PCR here were the same as those used for genomic PCRs.

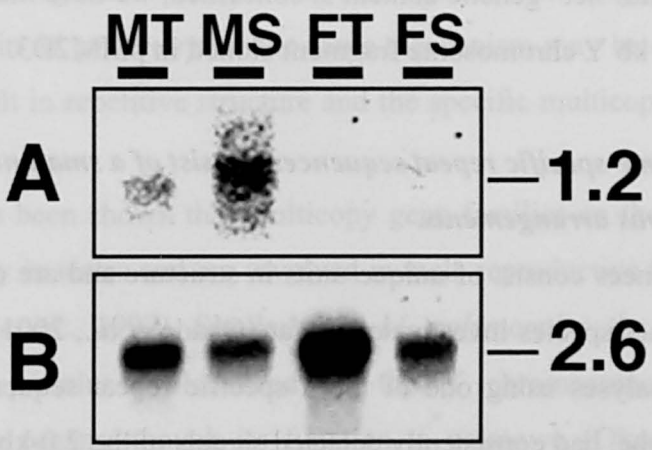


Fig. 1-8. Northern blot analysis of M2D3.4. (**A**) Five micrograms of poly(A)+RNA from male thalli (lane MT), male sex organs (lane MS), female thalli (lane FT), and female sex organ (lane FS) were blotted and probed with a [α-^{32}P]-labeled DNA covering M2D3.4. (**B**) The same membrane was reprobed with the CDPK gene, which is constitutively expressed in male and female. Sizes of signals are given in kb on the right.

DISCUSSION

In this study we provide comprehensive sequence information on approximately 90 kb of the liverwort Y chromosome. In addition to a member of the Y chromosome-specific gene family, ORF162, we have identified five novel genes whose copies have been amplified uniquely on the Y chromosome, though four of them also have homologues present on the X chromosome and/or autosomes.

The integrity of the pMM2D3 sequence

Repeat sequences have been always one of the major problems in assembling sequence reads or genomic clones into the correct order. "Complete" genome sequences of eukaryotes generally contain gaps of highly repetitive sequences (Mardis *et al.*, 2002). This is particularly true in the sex chromosomes in other species, as in *M. polymorpha*. For the region reported previously in pMM4G7, and now in pMM2D3, the Y chromosome-specific repeat sequences prevented us from reconstituting the sequences from these PAC clones. However, a copy of ORF162 embedded deeply within the repeat sequences was successfully identified, demonstrating that it is feasible to uncover genes in highly repeated sequences without reconstituting the precise primary structure of the surrounding repetitive regions, provided that the non-repetitive region is fully represented by shotgun sequence reads. As far as the "net" genetic content is concerned, we have indeed revealed the genetic content of the 90 kb Y chromosome fragment cloned in pMM2D3.

The Y chromosome-specific repeat sequences consist of a small number of repeat elements in various arrangements.

The repeat sequences consist of unique units in structure and are distinct in their sequence from those in other species investigated so far (Okada *et al.*, 2001). The previous genomic Southern blot analyses using one of the Y-specific repeat sequences, the 2.4-kb *Bam*HI fragment as a probe, had consistently detected signals of the 2.0-kb and 1.8-kb fragments in the male genomic DNA (Okada *et al.*, 2001), confirming that the two novel 2.0-kb and 1.8-kb *Bam*HI fragments found in pMM2D3 are also major repeat arrangements accumulated in the liverwort Y chromosome.

Multicopy genes on the Y chromosome

The five multicopy genes newly identified in this study seem to have one or more homologous sequences on autosomes, but a possibility that those homologous sequences in the female genome are present in the same restriction context on the X chromosome cannot be excluded. In either case, it is clear that these five genes have been amplified in the Y chromosome. Therefore, it is now evident that not only non-coding repeat sequences, but also embedded genes in the repeat sequences contribute to the repetitive nature of the liverwort Y chromosome. Our preliminary estimation by dot blot analysis suggests that at least a few dozens of copies of these genes have accumulated in the male genome (data not shown). Consistently, partial cDNA sequencing also revealed the presence of other copies of the genes with sequence variation (data not shown).

It is well established that the lack of recombination in most of the Y chromosome results in the accumulation of repeat sequences as well as of mutations (Charlesworth *et al.*, 1991). Y chromosomes have been reported to harbor unique repeat sequences in various organisms (Shibata *et al.*, 1999; Okada *et al.*, 2001; Foote *et al.*, 1992; Devlin *et al.*, 1998). A recent study reported that tandem duplication and inversion of large sequence blocks of 115-678 kb, which contain active genes, resulted in an extensive palindromic complex and simultaneous amplification of the transcriptionally active genes in a portion of the human Y chromosome (Kuroda-Kawaguchi *et al.*, 2001). Although no such long repeat units have thus far been identified in *M. polymorpha,* the same mechanism may be operating also in *M. polymorpha* and result in repetitive structure and the specific multicopy genes of the Y chromosome.

In humans, it has been shown that multicopy gene families on the Y chromosome are expressed specifically in testes and are involved in male reproductive functions such as spermatogenesis (Lahn *et al.*, 1997). Similarly, in *M. polymorpha,* the ORF162 family encoding RING-finger proteins is localized on the Y chromosome and expressed specifically in male sex organs, though its function is unknown (Okada *et al.*, 2001). Similarly, transcripts of the multicopy genes represented by M2D3.4 are detectable by northern blot analysis in male sex organs at much higher level than in male thalli (Fig. 1-8). On the other hand, there is no detectable mRNA in female thalli and sex organs (Fig. 1-8). Given its similarity to a protein localized in male gametic cells in lily (Xu *et al.*, 1999), the

M2D3.4 gene product may have a function in the male reproductive system in *M. polymorpha*.

On the other hand, the other three gene families, M2D3.1, M2D3.2 and M2D3.6, are expressed both in male and female constitutively, indicating that these three genes most likely do not have sex-specific functions. Genes that are most likely house-keeping function are also accumulated in the liverwort Y chromosome. The expression of these genes on female indicates that the homologues on autosomes and the X chromosome are active and the copies on the Y chromosome possibly are not expressed. In a case that the active homologue is on autosomes, the copies on the Y chromosome is not necessary and these should be in the process of degeneration and eventually elimination. In the other case that the active homologue is on the X chromosome, the copies on the Y chromosome is also necessary and at least one of them should remain to be active. The investigation of the homologues on autosome and/or the X chromosome makes us possible further insight to the evolution of the gene families.

Three molecular evolutionary processes are known to contribute to the formation of multicopy gene families on the human Y chromosome: transposition of autosomal transcription units (such as *DAZ*) (Saxena *et al.*, 1996), retroposition of autosomally encoded mRNAs (*CDY*) (Lahn and Page, 1999a) and persistence of genes previously shared with the X chromosome (*RBMY*) (Delbridge *et al.*, 1999; Mazeyrat *et al.*, 1999). For example, four copies of the *DAZ* gene arose during the primate evolution through the transposition and subsequent amplification of a single-copy autosomal gene, *DAZL*, that is still extant on human chromosome 3 (Saxena *et al.*, 1996; Saxena *et al.*, 2000).

Five gene families found in the liverwort Y chromosome have their homologues on autosomes or/and the X chromosome, suggesting that the same mechanism as in human, such as autosome-to-Y transposition or retrotransposition contributes to the evolution of the Y chromosome in *M. polymorpha*. For example, M2D3.1 has introns, which suggests the following two possibilities. One is that an autosomal copy of this gene transposed to the Y chromosome and amplified on the Y chromosome like the *DAZ* gene family in human (Saxena *et al.*, 1996). The other is that M2D3.1 homologue is on the X chromosome and M2D3.1 homologues on the Y chromosome and the homologues on the X chromosome evolved from a gene on the liverwort proto-Y and proto-X pair like the RBMY gene families in human (Delbridge *et al.*, 1999; Mazeyrat *et al.*, 1999).

Further investigation on these gene families, such as isolation of cDNA and transgenetic analyses, will reveal the evolution and function of the liverwort Y chromosome. The emerging facts on the liverwort Y chromosome should help us understand the function and evolution of plant sex chromosomes.

Chapter II

Sequence analysis of a 0.7-Mb region of the liverwort Y chromosome which does not carry the Y chromosome-specific repetitive sequences

INTRODUCTION

The liverwort Y chromosome has an approximately 4-Mb region which has accumulated the Y chromosome-specific repetitive sequences (Okada *et al.*, 2001). This region is designated as YR1, and the other region of the Y chromosome is designated as YR2 (approximately 6-Mb). 429 PAC clones, which contain the Y chromosome-specific repetitive sequences, have been isolated and are thought to cover most part of YR1 (Okada, 2002). A Y chromosome-specific gene, ORF162, was found in a PAC clone derived from YR1, and showed to be present in multiple copies, estimated as a few hundreds copies, on the Y chromosome (Okada *et al.*, 2001). In Chapter I, the author investigated another PAC clone derived from YR1 and identified five novel genes whose copies have been amplified, estimated as a few dozens copies, uniquely on the Y chromosome (Ishizaki *et al.*, 2002). Hence all the genes identified in YR1 to date are showed to be present in multiple copies on the Y chromosome.

As for YR2, nine Y chromosome-linked DNA fragments were isolated by representational difference analysis (RDA) (Fujisawa *et al.*, 2001; Yamamoto, 2002). These Y chromosome-linked DNA fragments were not carried by PAC clones covering most part of YR1, strongly suggesting that these DNA fragment were derived from YR2 (Hayashi, 2002; Yamamoto, 2002). In fact, a male PAC clone, pMM23-104E4, carries one of the Y chromosome-linked DNA fragments and was mapped on YR2 by FISH (Fujisawa, 2002). Additional FISH signals of pMM23-104E4 were detected also on the X chromosome and autosomes, indicating that YR2 contains not only the Y chromosome-specific sequences, but also sequences dispersed genome-wide. Sequence analysis of pMM23-104E4 indicated that retrotransposon-like sequences were one of the sequences shared by the male and female genomes in *M. polymorpha* (Fujisawa, 2002). Furthermore, a gene, *M104E4.1*, was identified in pMM23-104E4, and showed to be present in single copy and specific to the Y chromosome (Fujisawa, 2002).

This chapter describes sequence analysis of a 0.7-Mb region of YR2. Five novel genes were found in the region of YR2, suggesting difference in gene organization between YR1 and YR2.

MATERIALS AND METHODS

Plant materials

Thalli and sex organs of *M. polymorpha* (E lines) used are described in Chapter I.

Sequence analyses

Sequencing of six PAC clones, pMM23-408G1, pMM23-265H5, pMM23-420F5, pMM23-286B9, pMM23-578C3, and pMM23-88B7, was performed by shotgun method as previously described (Okada *et al.*, 2001). Draft sequence data of the six clones to 6-time coverage were provided by Shimadzu Biotech (Kyoto, Japan). A gap between pMM23-420F5 and pMM23-286B9 was filled by sequencing a bridge PCR fragment by primer walking. Sequence reads were assembled and edited by using computer programs Paracel Genome Assembler (PGA) (Paracel, Pasadena, CA, USA), PHRED (Ewing *et al.*, 1998), and CONSED version 12.0 (Gordon *et al.*, 1998). Sequence gaps were filled by primer walking. Searches for protein coding regions were performed against the non-redundant protein sequence database at the National Center for Biotechnology Information (NCBI) using BLASTX program (Altschul *et al.*, 1990) and against *M. polymorpha* ESTs (Nagai *et al.*, 1999, Nishiyama *et al.*, 2000) using BLASTN program (Altschul *et al.*, 1990). Amino acid sequences were aligned using CLUSTALW program (Thompson *et al.*, 1994).

Genomic PCR analysis

PCR for determining sex-specificity of putative genes was performed basically as described previously (Okada *et al.*, 2000). Template genomic DNAs were isolated as described by Takenaka *et al.* (2000), and 10 ng each was used as template. Sequences of primers are,
5'- CTGCTGATATCAATGCGCTTTT and
5'- ATCCGGCCTCCATTGTAATGGG for *M408G1.2*,
5'- ATAACAATGATTGCCGAAAGGGAT and
5'- ACCTGTAACACTGCTGCTTCACAC for *M286B9.1*,

5'- CCTCTTTACCATCCAACGTG and
5'- TTTGACACATTTGGCACAACTGAT for *M286B9.2*,
5'- AAACTTCATTGCTCCTCTCACA and
5'- AACTCGTCATCACCATACGCAC for *M578C3.1*,
5'- CTTGTCTATCGGTCATTCGCAT and
5'- GGTCCCTTGATTTCTTACAGCA for *M88B7.1*.

RT PCR analysis

Total RNAs from thalli and sexual organs of male and female plants were individually prepared as described in Chapter I. Poly(A)+ RNA was prepared with the PolyATract™ System 1000 (Promega) according to the manufacturer's instructions. First-strand cDNA was synthesized from 5 μg of DNase-treated poly(A)+ RNA using SuperScript II™ reverse transcriptase (Gibco BRL) at 42 °C with XhoSseEcoR-dT primer (5'- GAGAATTCCTGCAGGCTCGAGTTTTTTTTTTTTTTTT) for 60 min. A 20 μl reaction mixture was diluted to 400 μl with TE, and 1 μl of the diluted mixture was used as template in a 20 μl PCR amplification mix containing 10 pmol of the same primers used for the genomic PCR. Reactions without reverse transcriptase were performed to check genomic DNA contaminations.

Genomic Southern blot analysis

Five micrograms of total DNA were digested with appropriate restriction enzymes, and the resulting fragments were separated in a 1% agarose gel in 1x TAE buffer. After alkaline treatment and blotting onto a nylon membrane, hybridization was performed in a solution containing 5x Denhardt's reagent, 6x SSC pH 7.4, 0.5% SDS, 100 μg/ml denatured salmon sperm DNA, and 50% formamide at 42°C. To prepare probes, DNA fragments specific to the putative genes were amplified by PCR with appropriate PAC plasmid DNAs as templates and with the primer pairs described in genomic PCR analysis (MATERIALS AND METHODS). The PCR products were then labeled with [α-^{32}P] dCTP by another round of PCR. Membranes were washed for 1h in a solution containing 1x SSC and 0.1% SDS at 42°C followed by two washes with 0.1 x SSC and 0.5% SDS at 65°C for 1 hr. Radioactive signals were visualized with a BAS2000 Image Analyzer (Fuji Photo Film).

Rapid amplification of cDNA ends (RACE)-PCR of M88B7.1

For 3' RACE, first-strand cDNA was synthesized as described in the previous section, RT-PCR. PCR was performed with the female cDNA prepared above as a template by using gene-specific primers, M88B7.1-F (5'- CTTGTCTATCGGTCATTCGCAT) and M88B7.1-F2 (5'- ATGATGCGGAGAGGTCCATCTG) and XhoSseEco primer (5'-GAGAATTCCTGCAGGCTCGAGT).

For 5' RACE, synthesis of first strand cDNA and PCR were performed by using 5' RACE System for Rapid Amplification of cDNA Ends Version 2.0 (GIBCO BRL) according to the manufacturer's instruction, with gene specific primers, M88B7.1-R (5'- CCCACATCAGACCACAGGCATT), M88B7.1-R2 (5'- CTTGTTCGTGTGATGATGACCT), and M88B7.1-R3 (5'- GGTCCCTTGATTTCTTACAGCA).

RESULTS

Construction of a sequence-ready contig map of a region which contains two Y chromosome- linked DNA fragments, rgm6 and rsm62

A contig map was constructed across a portion of the Y chromosome using a combination of landmark content mapping and restriction digestion fingerprinting. First, two Y chromosome-linked DNA fragments, rsm62 and rgm6 (Fujisawa *et al.*, 2001; Yamamoto, 2002) were used as sequence tagged sites (STSs) to screen the gridded array of the male genomic PAC library (Okada *et al.*, 2000). Contigs extended from rsm62 and rgm6 were merged into a single contig of 970 kb which consists of 63 PAC clones (Fig. 2-1A). None of the 63 PAC clones were found among the 429 PAC clones from YR1 (Okada, 2002), suggesting that these clones represent part of YR2. In the process of aligning PAC clones, 37 additional male specific STSs were generated by PAC-end sequencing. Plasmid DNAs of isolated PAC clones were subjected to restriction digestion fingerprinting with *Bam*HI and *Not*I to establish extents of overlaps among them.

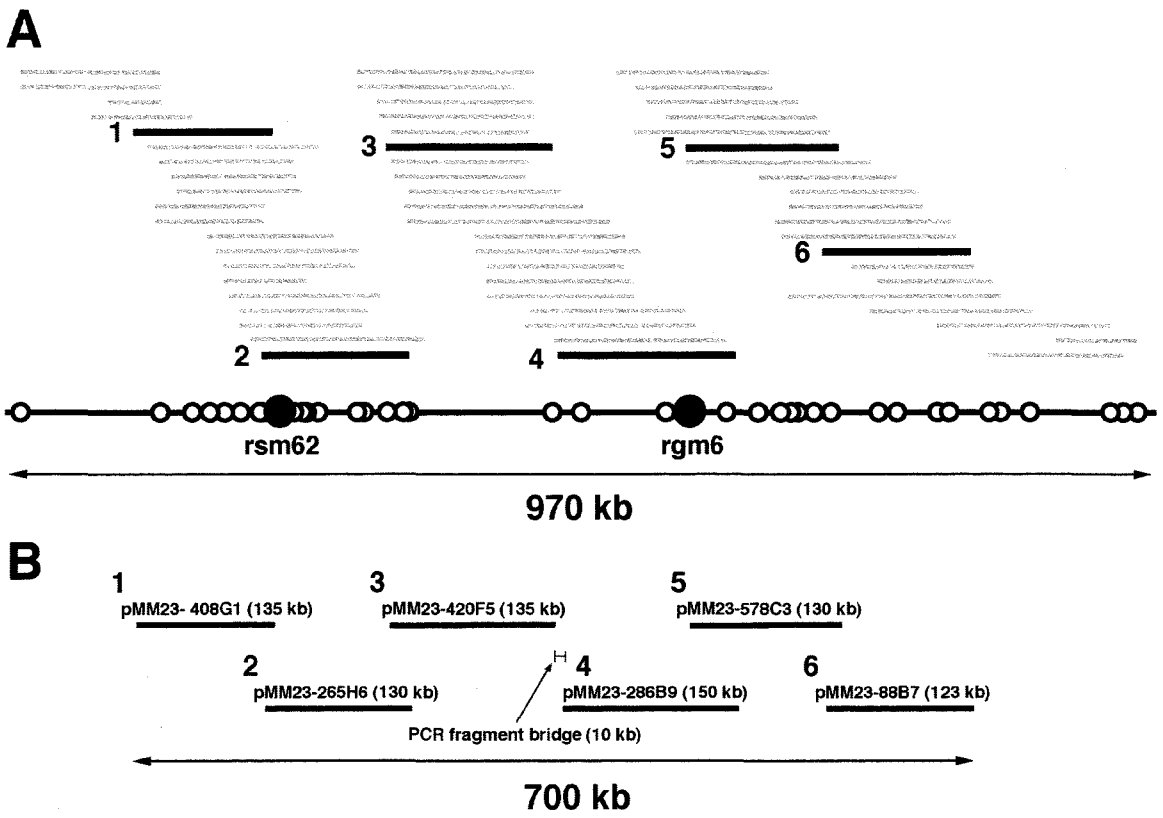


Fig. 2-1. (A) A contig map of the region containing rgm6 and rsm62. Horizontal bars represent PAC clones. Thick horizontal bars represent PAC clones chosen for sequencing. Closed circles indicate the Y chromosome-linked RDA markers, rsm62 and rgm6 (Fujisawa *et al.*, 2001). Open circles indicate male-specific STSs generated by PAC-end sequencing.
(B) PAC clones chosen for sequencing.

Sequence of a 0.7-Mb portion of the Y chromosome

Six PAC clones, pMM23-408G1, pMM23-265H5, pMM23-420F5, pMM23-286B9, pMM23-578C3, and pMM23-88B7, were selected out of the contig map and sequenced. A gap between two PAC clones, pMM23-420F5 and pMM23-286B9, was filled by sequencing a 10-kb PCR fragment which bridges the two clones (Fig. 2-1B). Total length of the region sequenced is 714,409 bp. No Y chromosome-specific repetitive sequence (Okada *et al.*, 2001) was found in the sequenced region, which is consistent with the result described in the previous section. This 0.7-Mb region carried sequences similar to various part of proteins encoded by genes in retrotransposons, *gag*, *pol*, and *int*. Retrotransposon-like sequences were found in sixty-one portion of the 0.7-Mb region (Fig.2-2).

Putative genes found in the sequenced region

Six regions, *M408G1.1*, *M408G1.2*, *M286B9.1*, *M286B9.2*, *M578C3.1*, and *M88B7.1*, show significant similarity to known proteins summarized in Table 2-1 at the deduced amino acid sequence level (E-value 1 x 10^{-10} or lower) and are depicted in Fig.2-2. The predicted reading frame of *M408G1.1* encodes four stop-codons in its frame, suggesting that this putative gene is a pseudo-gene.

Table 2-1 Potential genes found in the 0.7-Mb region

Sequence	% Identity (aa)	Similar sequences (species)	Accession No.
M408G1.1	52% (83/163)	T5J17.20 (*A. thaliana*)	T06091
M408G1.2	64% (678/1056)	putative protein phosphatase, T27D7.10 (*A. thaliana*)	AAF22889
M286B9.1	38% (111/287)	unknown protein (*A. thaliana*)	BAB10493
M286B9.2	42% (115/268)	bHLH protein like (*A. thaliana*)	NP_199488
	99% (422/427)*	male EST M01F020 (*M. polymorpha*)	AU081876
M578C3.1	57% (455/799)	product of EST AU082118 (*O. sativa*)	BAA89552
M88B7.1	55% (138/249)	putative flowering protein CONSTANS (*A. thaliana*)	AAK76580

* Identity as nucleotide sequence.

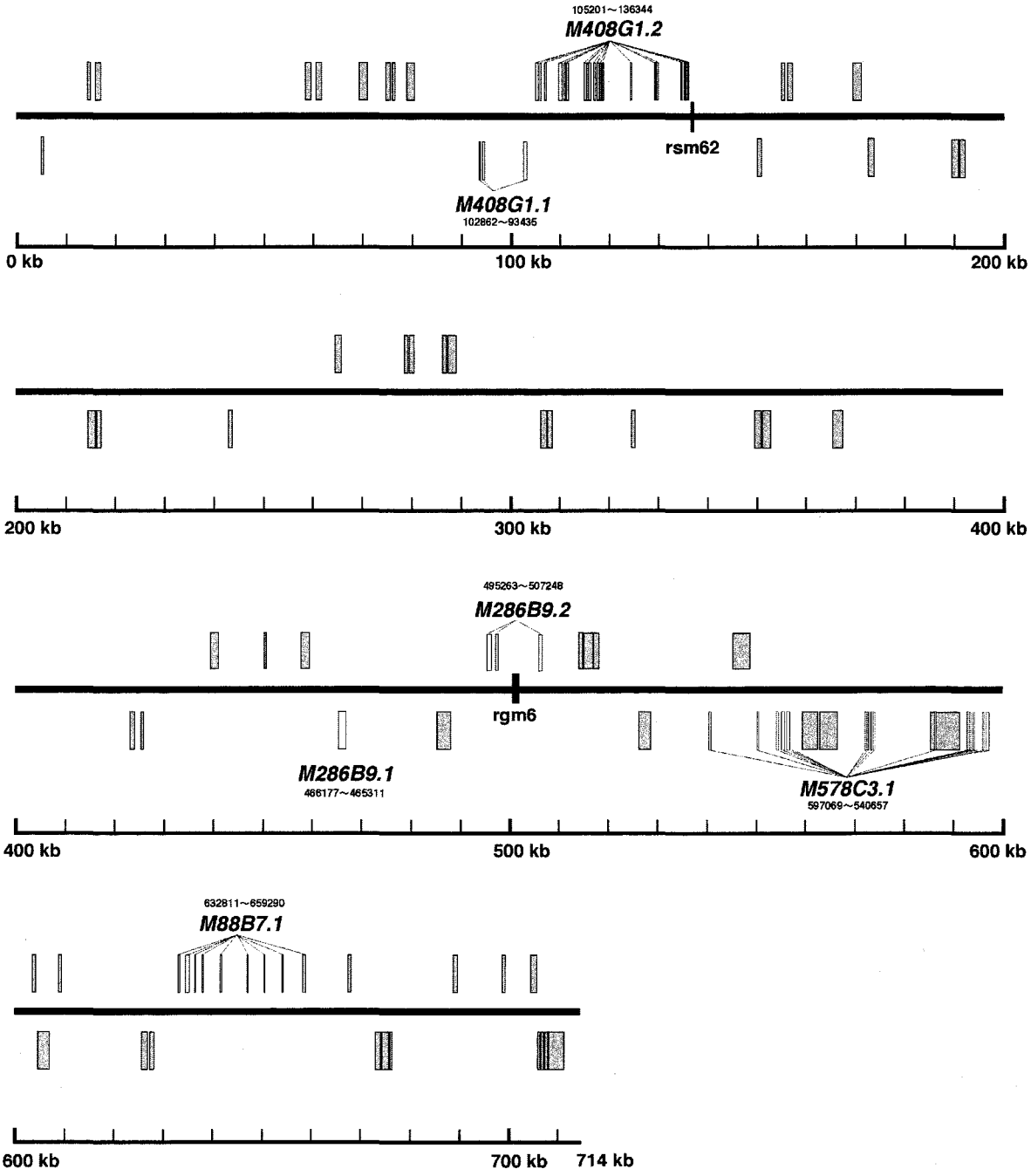


Fig. 2-2. Schematic diagram of the structure of a 0.7-Mb region sequenced. Putative genes are indicated by open boxes above (left-to-right orientation) and below (right-to-left orientation) the line. Regions which show similarity to genes found in retrotransposons at the amino acid sequence level are indicated by gray boxes. A closed box indicates the position of two Y chromosome-linked DNA fragments rsm62 and rgm6.

>M408G1.2

Although it was hard to deduce the entire structure of *M408G1.2* from the similarity to known proteins, the segmented alignment of amino acid sequences of *M408G1.2* and T27D7, probable protein of *A. thaliana*, suggested that *M408G1.2* consists of at least nineteen exons. The joined sequence of the nineteen putative exons contains an uninterrupted reading frame, and the dinucleotides GT and AG characteristic for 5'- and 3'-ends of eukaryotic introns are conserved at the boundaries of the deduced introns. Multiple alignment of the predicted product of *M408G1.2* and its related proteins revealed that *M408G1.2* contains a catalytic domain of protein phosphatase 2A (PP2A) (Fig. 2-3). PP2A is serine-threonine protein phosphatase in all eukaryotic cells, and it has been shown to regulate diverse cellular activities, including metabolism, DNA replication, transcription, RNA splicing, translation, cell cycle progression, morphogenesis, and development (Lechward *et al.*, 2001). The high degree of conservation of amino acid sequence between the predicted product of *M408G1.2* and probable protein of *A. thaliana* is not confined in the catalytic domain of PP2A (Fig. 2-3), suggesting their analogical functions.

>M286B9.2

M286B9.2 was tagged by one of ESTs from male sexual organs, indicating that the sequence is transcribed. By determining entire sequence of the EST clone, which tagged *M286B9.2*, partial coding sequence of it was determined. No stop codon was found in the 5' side of the cDNA sequence. The predicted product of *M286B9.2* contains a conserved basic-helix-loop-helix (bHLH) DNA binding domain as well as a nuclear localization signal, which was first identified in the basic region of Lc (Shieh *et al.*, 1993) (Fig. 2-4). A group of proteins, which contain a bHLH domain, are often referred as the myc homology family. An HLH region is required for dimerization, and an adjacent basic region is required for DNA binding (Pater *et al.*, 1997). The presence of a bHLH DNA binding domain suggests that the product of *M286B9.2* potentially binds to DNAs and regulates their transcriptional activity.

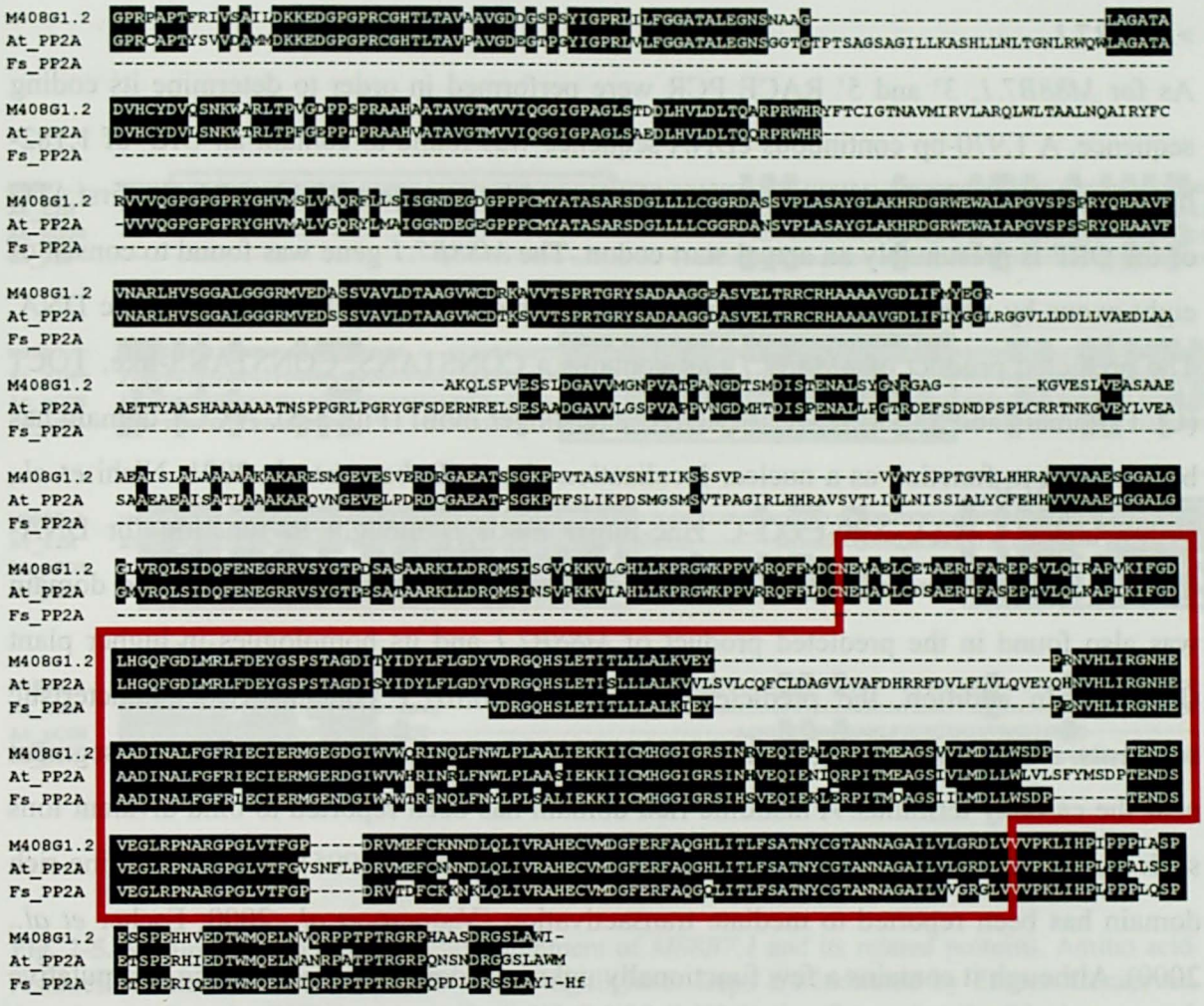


Fig. 2-3. Multiple amino acid sequence alignment of *M408G1.2* and its related proteins. Amino acid residues identical to those of *M408G1.2* are highlighted. Gaps are indicated by dashes. The PP2A, protein phosphatase 2A catalytic domain is indicated by red box. At_PP2A, probable protein of *A. thaliana* (GenBank accession no. AAF22889); Fs_PP2A, protein phosphatase of *Fagus sylvatica* (CAC11128).

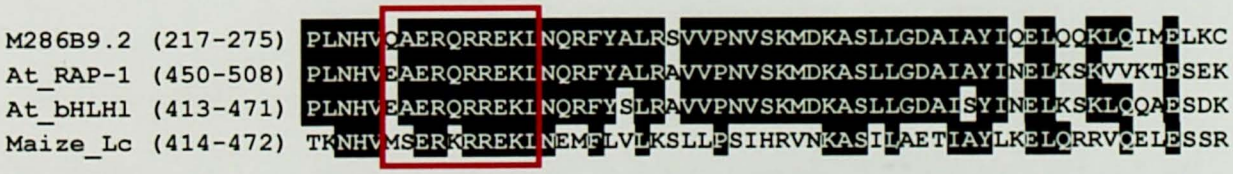


Fig. 2-4. Multiple amino acid sequence alignment of the bHLH domains found in *M286B9.2* and its related proteins. The numbers in parentheses indicate the positions of the domain in the respective sequences. Amino acid residues identical to those of *M286B9.2* are highlighted. The region homologous to one of the Lc nuclear localization signals is indicated by red box. At_RAP-1, RAP-1 of *A. thaliana* (NP_174541); At_bHLH1, bHLH protein-like of *A. thaliana* (NP_199488); Maize_Lc, regulatory protein Lc of *Zea mays* (M26227).

>M88B7.1

As for *M88B7.1*, 3' and 5' RACE PCR were performed in order to determine its coding sequence. A 1,970-bp continuous cDNA sequence was found to contain an ORF of 1,182-bp. Because an in-frame stop codon was present at 6 bp upstream of this ORF, the first ATG of the ORF is presumably an actual start codon. The *M88B7.1* gene was found to consist of eight exons by comparison of the nucleotide sequences of the cDNA and genomic DNA. The predicted product of *M88B7.1* and contains a CONSTANS, CONSTANS-like, TOC1 (CCT) domain and a C-X2-C-X20-C-X2-C zinc-finger motif (Fig. 2-5). A CCT domain has been shown to function as a nuclear localization signal (Robson *et al.*, 2001; Nishi *et al.*, 2000), and a C-X2-C-X20-C-X2-C zinc-finger motif is thought to function for DNA-binding, transcription activation (Nishii *et al.*, 2000). An unnamed but conserved domain was also found in the predicted product of *M88B7.1* and its homologues in higher plant (Fig.2-5). In addition, the predicted product of *M88B7.1* contains two characteristic segments: a histidine-rich segment near the amino terminus, and a glutamine-rich segment near the carboxy terminus. A histidine-rich domain has been reported to bind divalent ions such as, Cu, Zn, and Ni (Battistoni *et al.*, 2001; Gilbert *et al.*, 1995), and a glutamine-rich domain has been reported to mediate transactivation (Vaquero *et al.*, 2000; Escher *et al.*, 2000). Although it contains a few functionally unknown domains, the presence of a putative nuclear localization domain and a potential DNA-binding, transcription activation domain, indicates the product of *M88B7.1* is a transcription regulator.

Fig. 2-5. Multiple amino acid sequence alignment of *M88B7.1* and its related proteins. Amino acid residues identical to those of *M88B7.1* are highlighted. Gaps are indicated by dashes. Conserved domains are boxed : CCT domain (red), C-X2-C-X20-C-X2-C zinc-finger domain (blue), unnamed domain (green), a histidine-rich and a glutamine-rich segments (purple). respectively. Arrowheads above indicate positions of cysteine residues of the putative zinc finger. At_ZIM, ZIM protein of *A. thaliana* (BAA97678); At_pCON, putative flowering protein CONSTANS of *A. thaliana* (AAK76580); Os_pZF, putative zinc-finger protein of *Oryza sativa* (AAK14414).

The five putative genes are male-specific and present in single-copy

To investigate their sex specificities of the five putative genes *M408G1.2*, *M286B9.1*, *M286B9.2*, *M578C3.1*, and *M88B7.1*, a primer pair was designed for the conserved region of each of the putative genes, and diagnostic genomic PCR was performed (Fig. 2-6A). All the five putative genes yielded a PCR product for male genomic DNA but not for female genomic DNA, indicating that the five putative genes are male-specific.

In order to verify the male-specificity of the five putative genes and to investigate their copy numbers, genomic Southern blot analysis was performed (Fig. 2-6B). In all cases, only one signal was detected in male DNA but not in female DNA at the size expected based on genomic sequence, indicating that all the five genes are single-copy genes unique to the Y chromosome.

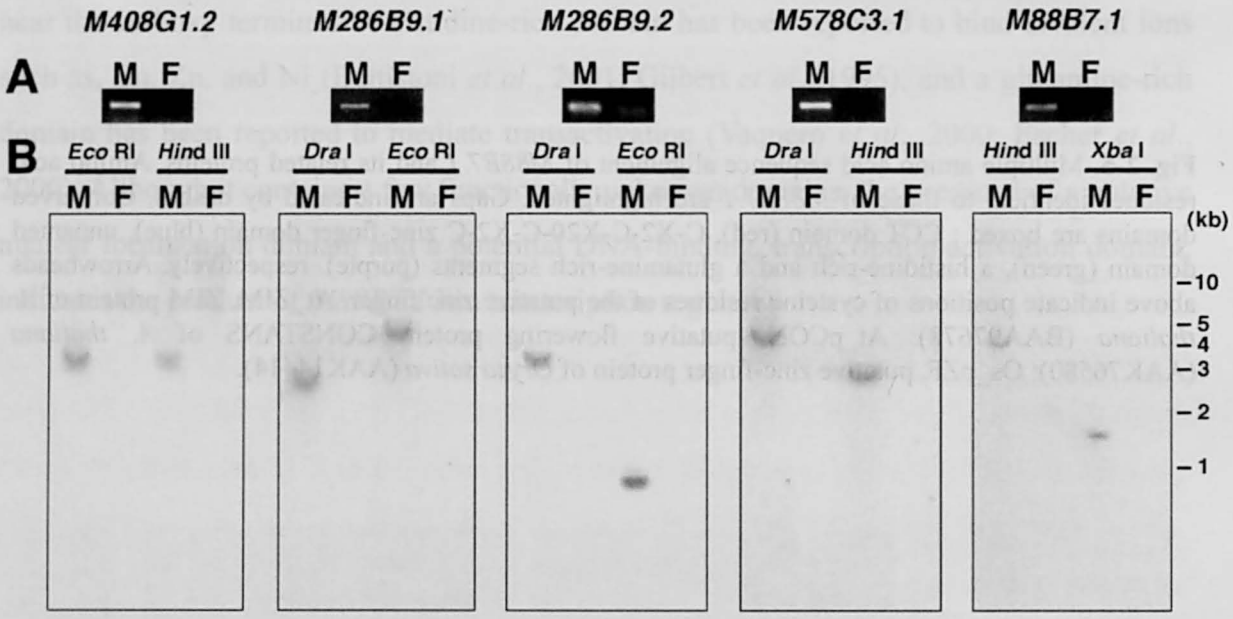


Fig. 2-6. The male-specificity and copy number of the five putative genes. **(A)** Male-specificity of the five genes examined by genomic PCR using genomic DNAs of male (M) and female (F) plants as a template and with primer pairs designed for the conserved regions of the respective genes. **(B)** Genomic DNAs of male (M) and female (F) plants were digested with two restriction enzymes, and probed with DNA fragments of the respective genes. Size standards are given in kb on the right.

The five putative genes are transcribed both in thalli and sex organs

Diagnostic RT-PCR was carried out to detect transcript of the five putative genes, *M408G1.2*, *M286B9.1*, *M286B9.2*, *M578C3.1*, and *M88B7.1*. All the five putative genes were shown to be transcribed in thalli as well as in mature sex organs (Fig. 2-7).

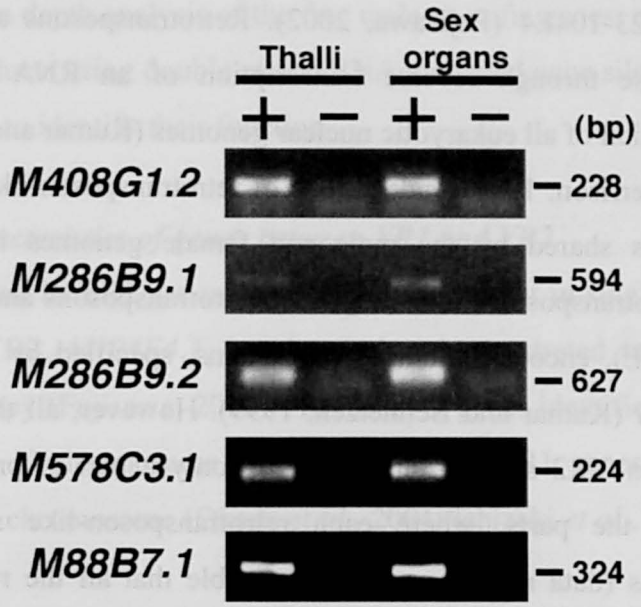


Fig. 2-7. Expression of the five putative genes examined by reverse transcription (RT)-PCR. Poly(A)+ RNAs isolated from male thalli and sex organs were reverse-transcribed and used as PCR templates (lanes +). The same amounts of poly(A)+ RNA with no RT reaction were used as negative control (lanes –). The primer pairs used for the genomic PCR shown in Fig. 2-5A were used. Sizes of PCR products are given in bp on the right.

DISCUSSION

In this chapter the author provided sequence information on 714,409 bp of YR2. The Y chromosome-linked DNA fragments isolated by RDA (Fujisawa *et al.*, 2001) made it possible to analyze such long continuous region of plant sex chromosome for the first time.

Sequence characteristics of the part of YR2

Retrotransposon-like sequences are dispersed in the 0.7-Mb region like another 134-kb region of YR2, pMM23-104E4 (Fujisawa, 2002). Retrotransposons are mobile genetic elements that transpose through reverse transcription of an RNA intermediate, and constitutes a major portion of all eukaryotic nuclear genomes (Kumar and Bennetzen, 1999; Schmidt and Heslop-Herrison, 1998). Indeed, one of retrotransposon-like sequences found in pMM23-104E4 was shared by the male and female genomes in *M. polymorpha* (Fujisawa, 2002). Retrotransposons, including LTR retrotransposons and long interspersed nuclear elements (LINE), encode a number of proteins, specified by three major genes called *gag*, *pol*, and *int* (Kumar and Bennetzen, 1999). However, all the retrotransposon-like sequences found in YR2 showed similarities to only a part of proteins encoded by retrotransposons, and the parts, where each retrotransposon-like sequences showed similarity, were various (data not shown). It is possible that all the retrotransposon-like sequences found in YR2 are not intact. In contrast to YR2, no retrotransposon-like sequence has been found in YR1 PAC clones sequenced to date (Okada *et al.*, 2001; Ishizaki *et al.*, 2002), suggesting that dispersion of retrotransposon-like sequences differs in YR1 and YR2.

Dispersion of male-specific and Y chromosome-linked sequences in the part of YR2 (Fig. 2-1A) indicates that recombination of the part of YR2 has been suppressed.

The five genes found in YR2

Five genes and one pseudo-gene were found in the 0.7-Mb region by similarity search. In *A. thaliana*, the only plant of which entire genome has been sequenced, 25,498 genes are encoded in the entire genome, approximately 120-Mb, and the gene density is estimated as 4~5-kb per gene (The Arabidopsis Genome Initiative, 2000). The total number of genes in *M. polymorpha* must not be so different from *Arabidopsis thaliana*, and the genome size of the *M. polymorpha* is estimated as ~280-Mb (Sone, 1999). Hence the gene density of *M.*

polymorpha is expected to be 10~20-kb per gene, suggesting further genes in the 0.7-Mb region sequenced in *M. polymorpha*.

The pseudo-gene, *M408G1.1*, might once be a protein-coding gene, and have degenerated on the Y chromosome since this gene turned into unnecessary. There is possibility of active homologous gene of *M408G1.1* on autosomes or the X chromosome. The potential products of the five male-specific genes, some of which showed similarity to transcriptional regulators or signal transducers, are possibly involved in male-specific functions. Further in depth analysis of the five male-specific genes, e.g. suppression of their transcriptional products using double-strand RNA mediated gene silencing (Kajikawa *et al.*, 2003), are required to identify their function.

Difference in characteristics of genes between YR1 and YR2

The five genes found in YR2 are male-specific and present in single-copy. Another male-specific gene in YR2, *M104E4.1*, has been also demonstrated to be male-specific and present in single-copy (Fujisawa, 2002). All the YR2 genes identified to date are present in single-copy on the Y chromosome, contrasting to six YR1 genes, which are present in multicopy on the Y chromosome (Okada *et al.*, 2001; Ishizaki *et al.*, 2002) (Fig. 2-8).

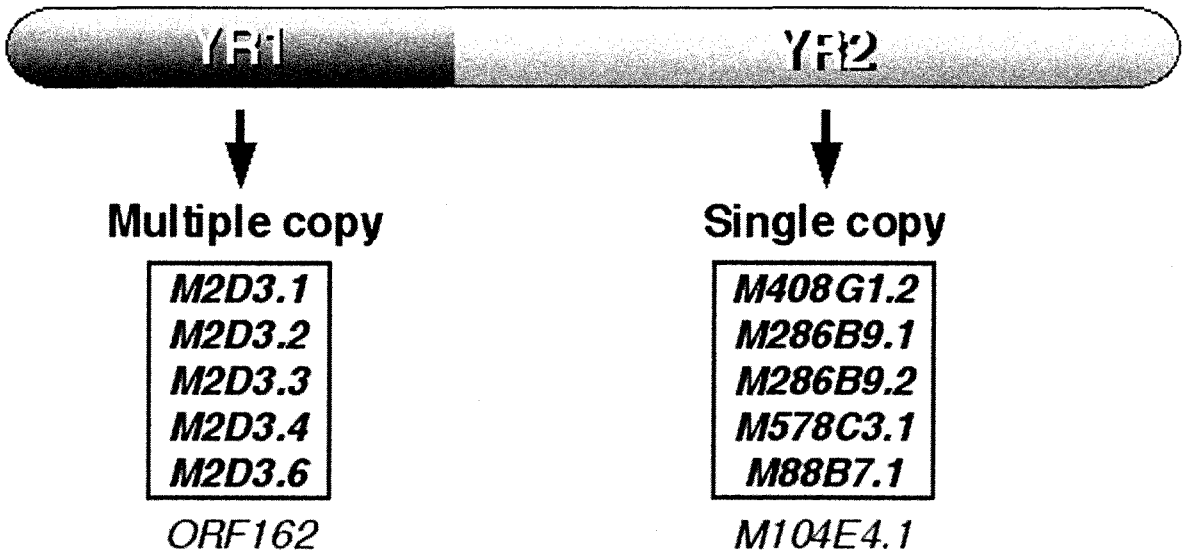


Fig. 2-8. Gene organization of the Y chromosome of the liverwort, *M. polymorpha*. All genes identified in YR1 to date are present in multiple copy in the Y chromosome. All genes found in YR2 to date are present in single copy and specific to the Y chromosome.

To the author's knowledge, this kind of difference in the copy number in respect to their locations in a chromosome has not been reported for plant genes.

In human, all of the genes identified in the AZFc (azoospermia factor c) region of the Y chromosome belong to multicopy gene families and seem to be expressed predominantly or exclusively in testis (Kuroda-Kawaguchi *et al.*, 2001), whereas roughly half of the genes on the non recombining region Y (NRY) are single copy genes, and are ubiquitously expressed (Lahn *et al.*, 2001, Bachtrog and Charlesworth, 2001). The five genes found in YR2 are transcribed both in thalli and sex organs. Another male-specific gene in YR2, *M104E4.1*, has been also showed to be transcribed both in thalli and sex organs (Fujisawa, 2002). All the YR2 genes identified to date are transcribed both in thalli and sex organs, contrasting to two genes in YR1 are transcribed specifically or predominantly in male sex organs (Okada *et al.*, 2001; Chapter I). It is also possible in *M. polymorpha* that YR1 and YR2 differ not only in structure and gene organization but also in gene expression pattern.

Chapter III

Comparison of X and Y chromosomal genes in the liverwort, *Marchantia polymorpha*

INTRODUCTION

The Y and X chromosomes of *M. polymorpha* contain clusters of their own repetitive sequences. The Y chromosome-specific repetitive sequences are accumulated in a 3~4-Mb region of the Y chromosome (Okada *et al.*, 2001), and thus the Y chromosome is divided into YR1, which is rich in Y chromosome-specific repetitive sequences, and YR2. Unlike the Y chromosome, ribosomal RNA genes (rDNAs) are accumulated on the X chromosome but not on the Y chromosome (Nakayama *et al.*, 2001). However the X chromosome can also be divided into two distinct segments: the rDNA cluster, and the rest. The other region of the X chromosome appears to contain sequences shared by YR2 (Fujisawa, 2002), but its sequence content remained mostly unclear.

Two X chromosome-linked DNA fragments have been isolated by representational difference analysis (RDA), one contains a 5S rRNA gene, and the other matches with two expressed sequence tags (ESTs) from immature female sex organs (Nagai *et al.*, 1999), F01I154 and F01Q066 (GenBank accession numbers C96067 and C96366, respectively). The one, which matches ESTs, is most likely derived from the non-rDNA portion of the X chromosome.

A gene, *M104E4.1*, found in YR2 has been indicated to be present in single-copy in YR2, and contain a LOV (light, oxygen, and voltage) domain, which has been found in blue light receptors of plant (Fujisawa, 2002). In the previous chapter, the author found another five single-copy genes in YR2. Although no homologue of YR2 genes has been detected in the X chromosome by genomic Southern blot analysis, it is still possible that there are X chromosome-linked degenerate homologues of the Y chromosomal genes in *M. polymorpha*. In human NRY (non recombining region Y), many of the Y chromosome-linked genes have their own degenerate homologues on the X chromosome (Lahn and Page, 1999b).

In this chapter the author describes sequence analysis of a PAC clone derived from non-rDNA portion of the X chromosome. A novel X chromosomal gene, *F62B12.1*,

containing a LOV domain was found in the PAC clone, and similar to a Y chromosomal gene, *M104E4.1*, at amino acid sequence level. This finding provides us an opportunity to understand the evolution of the X- and Y-chromosomal genes in *M. polymorpha*.

MATERIALS AND METHODS

Plant materials

Thalli and sex organs of *M. polymorpha* (E lines) described in Chapter I were used.

Sequence analyses

Shotgun sequencing of pMF28-62B12 was performed as described previously (Okada *et al.*, 2000). Draft sequence data of pMF28-62B12 to 6-time coverage was provided by Shimadzu Biotech (Kyoto, Japan). Searches for protein coding regions were performed against the non-redundant protein sequence database at the National Center for Biotechnology Information (NCBI) using BLASTX program (Altschul *et al.*, 1990) and against *M. polymorpha* ESTs (Nagai *et al.*, 1999; Nishiyama *et al.*, 2000) using BLASTN program (Altschul *et al.*, 1990). Amino acid sequences of *F62B12.1* and *M104E4.1* were aligned using CLUSTALW program (Thompson *et al.*, 1994).

RNA preparation and Northern blot analysis

Total RNAs from male and female thalli, male and female sexual organs were individually prepared as described in Chapter I. Poly(A)+ RNA was prepared with the PolyATract™ System 1000 (Promega) according to the manufacturer's instructions. Five micrograms of poly(A)+ RNA were electrophoresed in a 0.8% denaturing agarose gel containing formaldehyde and transferred onto a nylon membrane. To prepare probe for *F62B12.1*, a DNA fragment specific to the gene was amplified by PCR with PAC plasmid DNA of pMF28-62B12 as a template and with the primer pair, F62B12.1-F and F62B12.1-R. Similarly, probe for *M104E4.1* was amplified by PCR with PAC plasmid DNA of pMM23-104E4 as a template and with the primer pair, M104E4.1-F and M104E4.1-R, (described in MATERIALS AND METHODS of Chapter II in Fujisawa, 2002). The PCR products of *F62B12.1* and *M104E4.1* were labeled with [α-^{32}P] dCTP by another round of PCR. Hybridization was performed with ExpressHyb™ solution (CLONTECH) as described (Okada *et al.*, 2001). The hybridized membrane was washed for 1 hr in a solution

containing 2 x SSC and 0.1% SDS at 25°C followed by two washes with 0.1 x SSC and 0.5% SDS at 55°C for 1hr. Radioactive signals were visualized with a BAS2000 Image Analyzer (Fuji Photo Film).

Rapid amplification of cDNA ends (RACE)-PCR

For 3' RACE, first-strand cDNA was synthesized from 1 μg of DNase-treated poly(A)+ RNA using SuperScript II™ reverse transcriptase (Gibco BRL) at 42°C with XhoSseEcoR-dT primer (5'- GAGAATTCCTGCAGGCTCGAGTTTTTTTTTTTTTTTT) for 60 min. PCR was performed with the cDNA prepared above as a template and gene-specific primers, F62B12.1-F (5'-CGATGAGGCTCTTTCTGAGAGA) and F62B12.1-F2 (5'-ACTTCTGAATTTATGCGGGTTG), and XhoSseEco primer (5'-GAGAATTCCTGCAGGCTCGAGT).
For 5' RACE, synthesis of first strand cDNA and PCR were performed by using the 5' RACE System for Rapid Amplification of cDNA Ends Version 2.0 (GIBCO BRL) according to the manufacturer's instructions, with gene specific primers, F62B12.1-R (5'-TTGGCCGTATAACCAGTCATAC), F62B12.1-R2 (5'-CGATGAGGCTCTTTCTGAGAGA), and F62B12-R3 (5'-ATGGAAAGATGCTCTTCAATGC).

Genomic Southern blot analysis

Five micrograms of genomic DNA were digested with *Pvu*II and the resulting fragments were separated in a 0.8% agarose gel in 1 x TAE buffer. After alkaline treatment and blotting onto a nylon membrane, hybridization was performed in a solution containing 5 x Denhardt's reagent, 6 x SSC pH 7.4, 0.5% SDS, 100 μg/ml denatured salmon sperm DNA, and 50% formamide at 42°C. To prepare probes, DNA fragment specific to the gene, *F62B12.1* was amplified by PCR with plasmid DNA of pMF28-62B12 as template and with the primer pairs, F62B12.1-F and F62B12.1-R, and then the PCR products were labeled with [α-^{32}P] dCTP by another round of PCR. Membranes were washed for 1hr in a solution containing 1 x SSC and 0.1% SDS at 42°C followed by two washes with 0.1 x SSC and 0.1% SDS at 65°C for 1 hr. Radioactive signals were visualized with a BAS2000 Image Analyzer (Fuji Photo Film).

RESULT

Isolation of X-linked PAC clones

To examine the non-rDNA portion of the X chromosome, the female genomic library (Okada *et al.*, 2000) was screened by PCR using a primer pair for the X chromosome-linked DNA fragment, rhf73 (Fujisawa *et al.*, 2001). Four PAC clones were isolated and aligned into a single contig. The alignment of the four PAC clones was verified by their restriction profiles (data not shown), and by PCR using primer pairs for the end sequences of one of the PAC clones, pMF28-62B12 (Fig. 3-1). The end sequences, etF62B12 and esF62B12, were shown to be specific to the female genome by PCR (data not shown). These results indicate that the contig map of the four PAC clones represents a unique region of the X chromosome tagged by rhf73.

Sequence analysis of a PAC clone, pMF28-62B12

In order to characterize part of the non-rDNA portion of the X chromosome, one of the X-linked clones, pMF28-62B12, was selected and sequenced. The insert size of pMF28-62B12 was 111-kb (Fig. 3-2). Ten retrotransposon-like sequences were found in pMF28-62B12. These retrotransposon-like sequences are highly degenerate, suggesting that these retrotransposon-like sequences are not functional.

Putative genes found in pMF28-62B12

Two potential protein-coding regions were found in pMF28-62B12 as summarized in Table 3-1 and depicted in Fig. 3-2.

F62B12.1 showed similarity to a LOV1 (light, oxygen, and voltage) domain of phototropin1 proteins found in flowering plants, such as *A. thaliana* (Haula *et al.*, 1997), *Z. mays*, *O. sativa*, *Avena sativa* (Salomon *et al.*, 2001), fern, *Adiantum capillus-veneris*, and algae *Chlamydomonas reinhardtii* (Briggs *et al.*, 2001, and Kasahara *et al.*, 2002).

At the amino acid sequence level, *F62B12.1* also showed similarity to a Y chromosomal gene, *M104E4.1*, which also contains a LOV domain (Fujisawa, 2002), suggesting that the X chromosomal gene, *F62B12.1*, is a homologue of the Y chromosomal gene, *M104E1.1*. The partial similarity of *F62B12.1* and *M104E4.1* in their LOV domain is 73.5% at nucleotide sequence level. One of the *M. polymorpha* ESTs tags *F62B12.1*, indicating that this region is transcribed.

F62B12.2 has two stop codons in its frame, suggesting that this putative gene is not functional. No transcript of *F62B12.2* was detected in RT-PCR analysis using cDNA from thalli and mature sexual organs (data not shown).

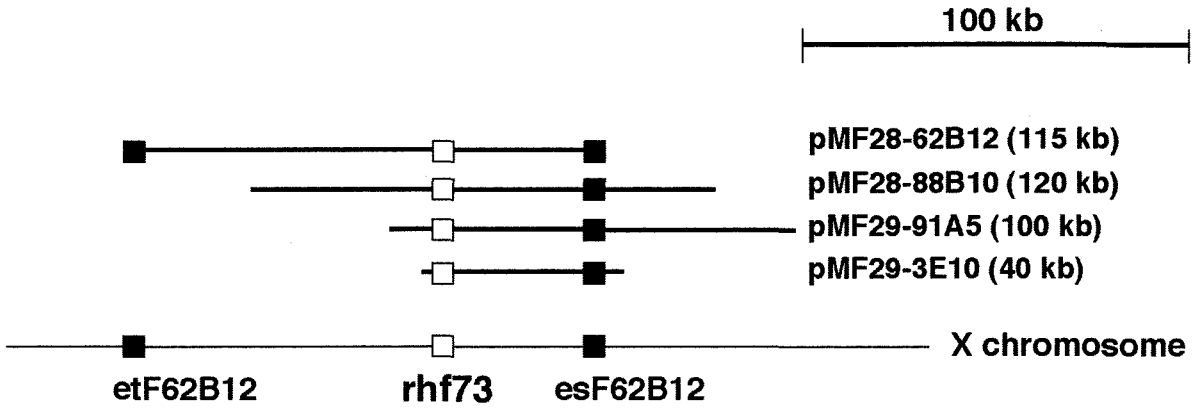


Fig. 3-1. A PAC contig map of an X chromosomal region containing rhm73. PAC clones are indicated by labeled horizontal lines. Squares indicate the presence of markers, rhf73, etF62B12, and esF62B12. These markers were detected by PCR using specific primer pairs. Primers for etF62B12 and esF62B12 were designed based on the SP6- and T7- end-sequences of pMF28-62B12, respectively. The exact position of rhm73 in each PAC clone and the location of the contigs on the X chromosome are undefined.

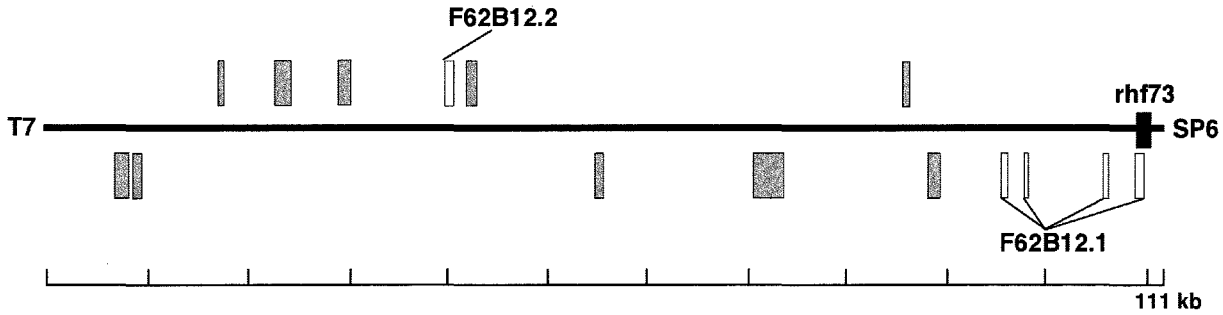


Fig. 3-2. Schematic diagram of the structure of pMF28-62B12. The insert of the plasmid is displayed with its SP6-end on the left. Putative genes are indicated by open boxes above (left-to-right orientation) and below (right-to-left orientation) the line. Gray boxes indicate regions which show similarity to genes found in retrotransposons at the amino acid sequence level. A closed box indicates the position of the X chromosome-linked DNA fragment, rhf73.

Table 3-1 Potential genes found in pMF28-62B12

Sequence	% Identity (aa)	Similar sequences (species)	Accession No.
F62B12.1	51% (151/355)	PHOT1 (*A. thaliana*)	AAC01753
	99% (797/801)*	Female sexual organ EST, F01105 (*M. polymorpha*)	C96067
	83% (88/107)	Male-specific gene, *M104E4.1* , LOV domain (*M. polymorpha*)	
F62B12.2	52% (111/215)	Putative protein kinase (*A. thaliana*)	NP_188973

*Identity as nucleotide sequence

F62B12.1 is female-specific and present in single copy

In order to investigate if the *F62B12.1* gene is specific to the X chromosome at the nucleotide sequence level, genomic Southern blot analysis was performed. Probe was designed to contain most part of the LOV domain. Probe for *F62B12.1* was hybridized only to female genomic DNA under the stringent condition described above. This is consistent with the observed degree of the similarity between the nucleotide sequences of *F62B12.1* and *M104E4.1*, a single-copy gene specific to the Y chromosome (Fujisawa, 2002). The size of the hybridization signal was 1.3 kb as predicted by a *Pvu*II restriction map of pMF28-62B12 (Fig. 3-3). These results indicate that *F62B12.1* appears to be a single-copy gene specific to the X chromosome but its divergent homologue is also present on the Y chromosome.

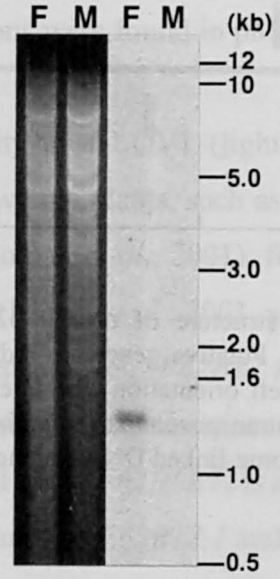


Fig. 3-3. Genomic Southern blot analysis of the *F62B12.1* gene. Female (lanes F) and male (lanes M) genomic DNAs were digested with *Pvu*II. Sizes of standard markers are given on the right.

Structure of the X chromosomal gene, F62B12.1

In order to determine the coding sequence of the *F62B12.1* gene, 3' and 5' RACE were performed. A continuous cDNA sequence of a 2,747-bp, which contains an ORF of 2,136 bp, was obtained (Fig. 3-4). Since an in-frame stop codon is present at 108 bp upstream of the ORF, and the nucleotide A in position –3 of the first ATG of the ORF is consistent to Kozak consensus sequence for initiation of translation (Kozak, 1989), the first ATG of the ORF is presumably a start codon. Comparison of the nucleotide sequences of cDNA and genomic DNA revealed that the *F62B12.1* gene consists of six exons (Fig. 3-4).

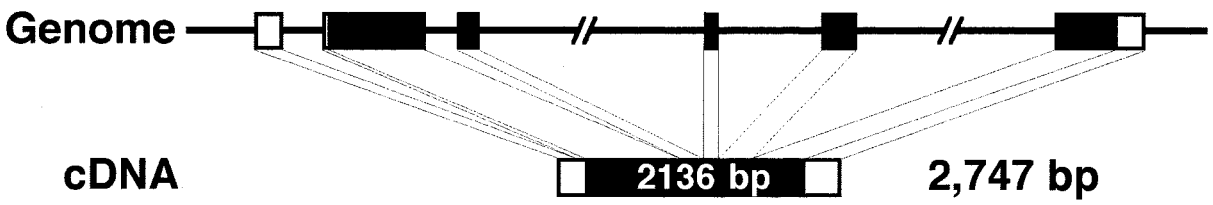


Fig. 3-4. Schematic representation for the structure of the *F62B12.1* gene and its cDNA. The 5'-end of *F62B12.1* is on the left. A horizontal line indicates the genomic sequence of the X chromosome. Open boxes indicate UTR regions, and closed boxes indicate putative coding sequences.

The transcript of *F62B12.1* potentially codes for 712 amino acid residues containing a well-conserved LOV domain which has been found in blue light receptors of plant, phototropins (Fig. 3-5B). All of the amino acid residues necessary for covalent binding with flavin mononucleotide (FMN) (Christie *et al.*, 1999; Salmon *et al.*, 2000) are conserved in the LOV domain of *F62B12.1* (Fig. 3-5B). No other known domains as motifs, such as a serine/threonine kinase domain carried by phototropin 1 (PHOT1) proteins (Fig. 3-5A),.was found in *F62B12.1*.

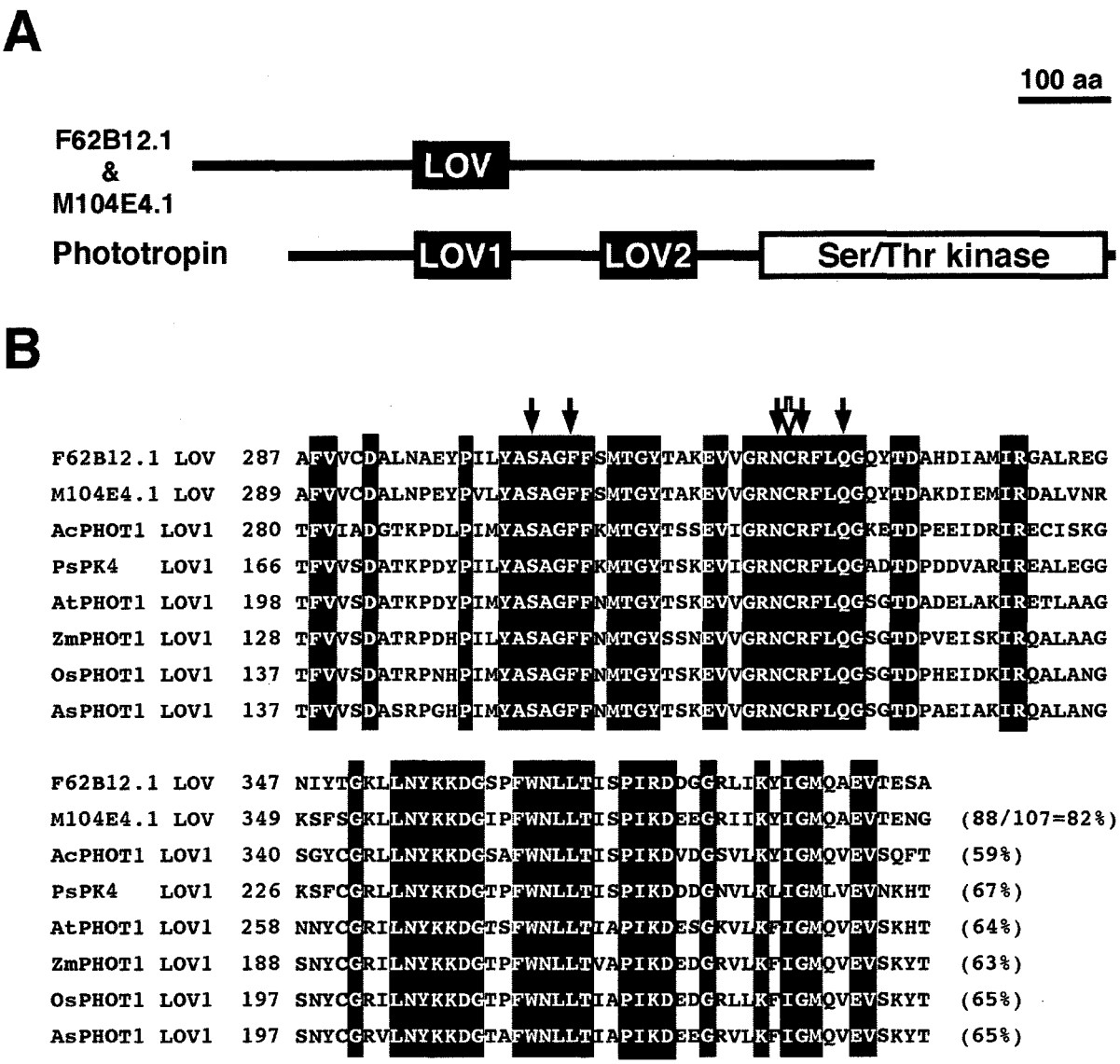


Fig. 3-5. Comparison of a putative protein encoded by the *F62B12.1* gene and related proteins. **(A)** Schematic representation of the structure of *F62B12.1* and *M104E4.1,* and phototropin1 proteins (PHOT1). Filled and open boxes indicate LOV domains and a serine/threonine kinase domain, respectively. **(B)** A local alignment of LOV domains conserved among *F62B12.1*, *M104E4.1*, and PHOT1 proteins is given with positions on the left. Phototropin1 proteins are from *Adiantum capillus-veneris* (AcPHOT1, accession number AAC05083), *Pisum sativum* (PsPK4, AAB41023), *A. thaliana* (AtPHOT1, AAC01753), *Z. mays* (ZmPHOT1, AAB88817), *O. sativa* (OsPHOT1, CAB65325), and *Avena sativa* (AsPHOT1, AAC05083). Amino acids conserved among all of the LOV domains are highlighted. Amino acid sequence identities with the LOV domain in *F62B12.1* are given in parentheses (identity% = the number of identical amino acid residues x 2 / sum of length of two sequences x 100). Filled arrows indicate amino acid residues required for holding a flavin mononucleotide (FMN), and an open arrow indicates the cysteine residue binding with FMN.

Comparison of the structures of the X- and Y-chromosomal genes in M. polymorpha

The deduced amino acid sequences of the X- and Y-chromosomal genes, *F62B12.1* and *M104E4.1*, were aligned (Fig. 3-6), and compared the lengths of their amino acid sequences are similar their global similarity is 44.4%, but two portions show remarkably high similarity, 91.0 % for the first 67 amino acid residues and 79.3 % in for the 193 amino acid residues including the LOV domain. Some other conserved but shorter stretches (less than 10 amino acid residues) were also found. All the four intron-insertion sites in the coding sequences are were completely identical between *F62B12.1* and *M104E4.1* (Fig. 3-6).

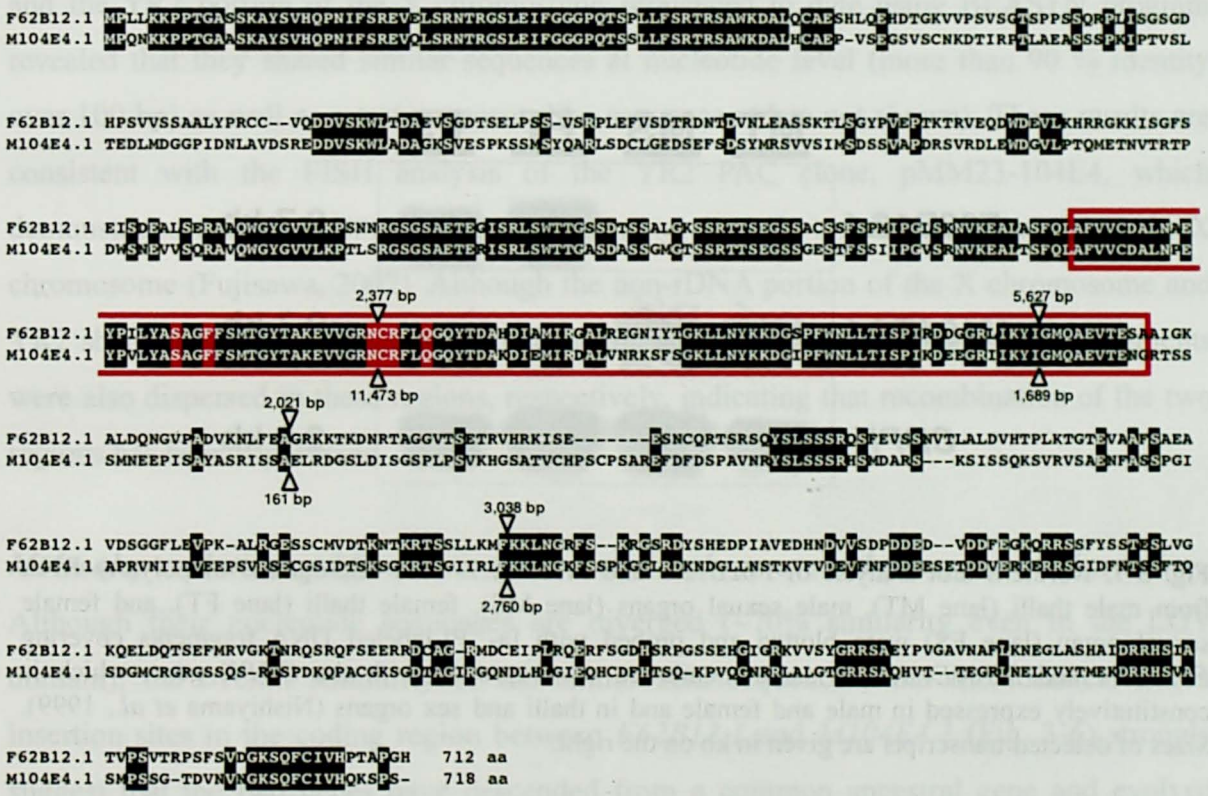


Fig. 3-6. Pairwise alignment of the deduced amino acid sequences of *F62B12.1* and *M104E4.1*. Amino acids conserved between *F62B12.1* and *M104E4.1* are highlighted. Alignment gaps are indicated by dashes. A LOV domain is boxed. Residues required for holding a flavin mononucleotide (FMN) are highlighted in red. Sites and sizes of introns are indicated by triangle and the numbers above or below the triangles respectively. Intron insertion sites and the numbers above or below besides the open triangles, indicates the sizes of the introns, respectively.

Expression of the sex chromosomal genes encoding a LOV domain.

In order to investigate expression pattern of the X- and Y-chromosomal genes containing a LOV domain, northern blot analysis of *F62B12.1* and *M104E4.1* was performed. A 2.7-kb transcript for *F62B12.1* was detected in female thalli and sex organs. The size of the transcript matches that of the *F62B12.1* cDNA. A transcript of 2.7-kb was detected for *M104E4.1* both in thalli and sex organs (Fig. 3-7). No cross hybridization between the two genes was observed under the experimental condition, as expected from the result of genomic Southern blot analysis.

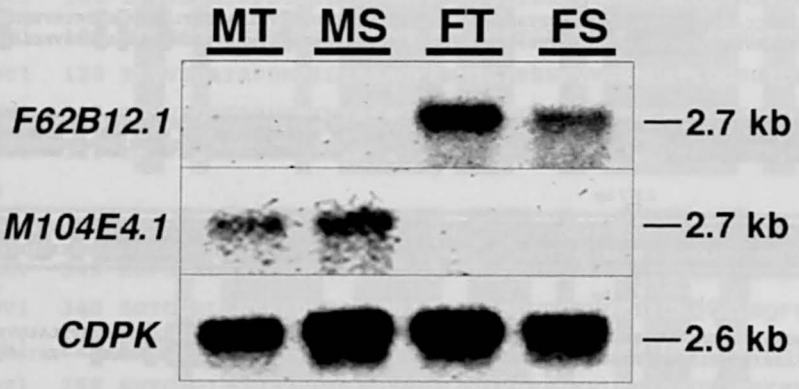


Fig. 3-7. Northern blot analysis of *F62B12.1* and *M104E4.1*. Five micrograms of poly(A)+RNA from male thalli (lane MT), male sexual organs (lane MS), female thalli (lane FT), and female sexual organ (lane FS) were blotted and probed with [a-^{32}P]-labeled DNA fragments covering *F62B12.1* and *M104E4.1*. The same membrane was reprobed with the *CDPK* gene, which is constitutively expressed in male and female and in thalli and sex organs (Nishiyama *et al.*, 1999). Sizes of detected transcripts are given in kb on the right.

DISCUSSION

In this chapter, the author determined and characterized 111-kb of the liverwort X chromosome, identifying a novel X-chromosomal gene, *F62B12.1* containing a LOV domain.

Comparison of sequence between the non-rDNA potion of the X chromosome and YR2

Sequence analysis of an X chromosome-derived PAC clone, pMF28-62B12, revealed that the non-rDNA portion of the X chromosome and YR2 of the Y chromosome share similar sequence characteristics, such as presence of dispersed retrotransposon-like sequences. Furthermore, sequence comparison of the 111-kb non-rDNA portion of the X chromosome and the YR2 portion of the Y chromosome sequenced to date using BLASTN program revealed that they shared similar sequences at nucleotide level (more than 90 % identity over 100 bp) as well as retrotransposon-like sequences (data not shown). These results are consistent with the FISH analysis of the YR2 PAC clone, pMM23-104E4, which demonstrated that YR2 contains sequence common not only to autosomes but also to the X chromosome (Fujisawa, 2002). Although the non-rDNA portion of the X chromosome and YR2 share similar sequences, X chromosome- and Y chromosome-linked DNA fragments were also dispersed in these regions, respectively, indicating that recombination of the two regions has been suppressed.

Molecular evolution and functions of the X and Y chromosomal genes for LOV proteins

Although their nucleotide sequences are diverged (~70% similarity even in the LOV domain), the overall similarity of the amino acid sequences, and the identical intron insertion sites in the coding region between *F62B12.1* and *M104E4.1* (Fig. 3-6) strongly suggest that the two genes have descended from a common ancestral gene and evolved independently in the X and Y chromosomes, respectively. In human, protein divergence of oldest X-Y gene pairs, *SOX3* and *SRY*, is 29% (Lahn and Page, 1999b), whereas protein divergence of the liverwort X-Y gene pair, *F62B12.1* and *M104E4.1*, is 54% (Fig. 3-6). The mutations fashioned one allele of *SOX3*, originally an autosomal gene, into the male-determining factor *SRY* (Lahn and Page, 1999b). It is also possible that the female- and male-specific functions are added to the X and Y chromosomal genes in *M. polymorpha*, it is also possible that the function of them have remained to be conserved. *F62B12.1* and

M104E4.1 are expressed both in thalli and sex organs, suggesting that they are involved in housekeeping functions (Fig. 3-7).

F62B12.1 and *M104E4.1* encode a protein having a domain that is most similar to a LOV1 domain of a blue-light receptor (Briggs *et al.*, 2001). Phototropins (PHOT1 and PHOT2, formerly designated NPH1 and NPH2) mediate phototropism (Haula *et al.*, 1997; Christie *et al.*, 1998), blue light-induced chloroplast relocation (Kagawa *et al.*, 2001; and Sakai *et al.*, 2001; Jarillo *et al.*, 2001), and blue-light induced stomatal opening (Kinoshita *et al.*, 2002) in *A.thaliana.* Other proteins which contain a LOV domain, ZTL (ZEITLUPE), LKP1 (LOV kelch protein 1), and LKP2 (LOV kelch protein 2), are dedicated to the light input pathway to the clock in *A. thaliana* (Somers *et al.*, 2000; Kiyosue *et al.*, 2000; Schultz *et al.*, 2001). A LOV domain itself has been demonstrated to bind a flavin mononucleotide and absorb blue-light (Salmon *et al.*, 2000). Conservation of amino acid residues characteristic to LOV domain (Fig. 3-5B) strongly suggests that the predicted products of *F62B12.1* and *M104E4.1* bind FMN and absorb blue-light. However, any of other proteins containing LOV domain, such as phototropins, has other functional domains. The two liverwort proteins containing a LOV domain may have novel functions involved in blue-light stimuli. Further functional analysis of *F62B12.1* and *M104E4.1* will clarify their functional difference and how it is involved in perception of light in *M. polymorpha.*

Evolution of sex chromosomal genes in the liverwort, M. polymorpha

Morphologically and genetically distinct sex chromosomes have arisen and evolved independently in many groups of animals and plants (Charlesworth B., 1991, Charlesworth D., 2002). Sex chromosomes are generally believed to have descended from a pair of ordinary autosomes. A pair of autosomes that would eventually become the sex chromosomes acquired a sex-determining role, and suppression of recombination between the nascent sex chromosomes allowed them to evolve independently (Charlesworth B., 1996).

In *M. polymorpha*, a specific Y chromosome is present in the male, and a distinct X chromosome is found only in the female plantlets. Since both X and Y chromosomes are present as haploid whether in zygotes or in gametes, these sex chromosomes may have been reciprocally accumulating unique sequences by suppression of recombination between

them. From the results of Chapter I and Chapter II, the author proposed a model that the copy number of genes differed with the region of the Y chromosome, YR1 and YR2. All the genes identified to date in YR1 are present in muticopy in the Y chromosome, whereas all the genes identified to date in YR2 are present in single-copy and specific to the Y chromosome. In Chapter III, the author found a degenerate X chromosomal homologue of a Y chromosomal gene, and they are suggested to have descended from a common ancestral gene.

As discussed in Chapter I, the genes found in YR1 may have amplified with long repeat units, resulting the repetitive structure of the YR1 and multicopy genes. In human, 4.5-Mb portion of the Y chromosome, that includes *AZFc* region, consists of complex blocks that repeated two or more times along the chromosome, and the blocks tend to be arranged in palindromes, or near-palindromes. Each block contains genes that are mirror images of each other.

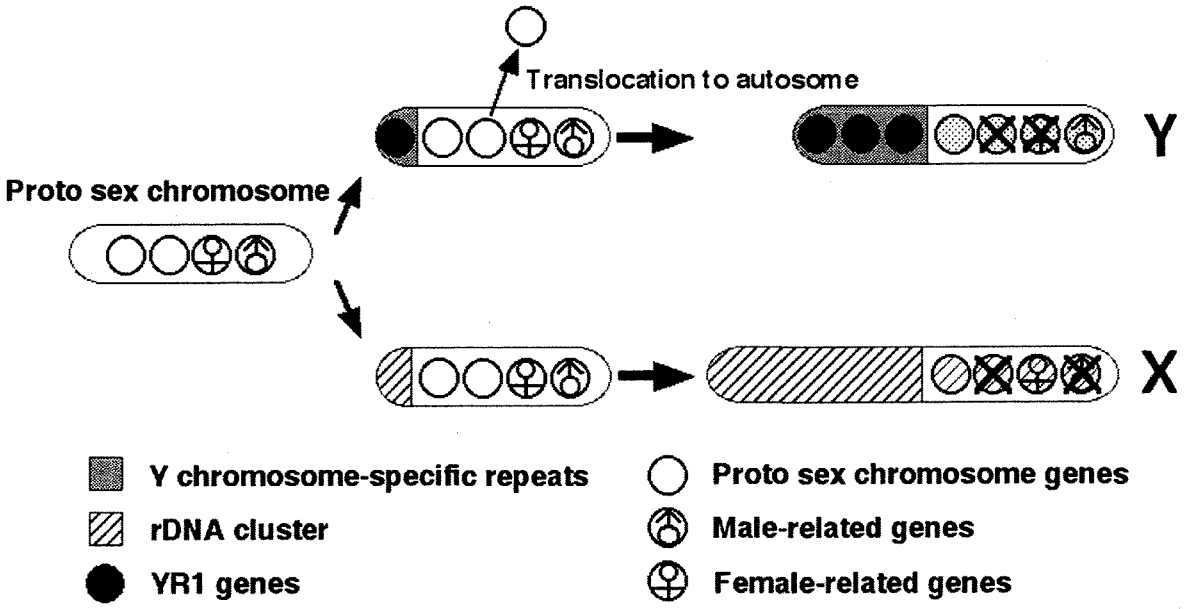


Fig. 3-8. A model of the evolution of sex chromosomal genes in *Marchantia polymorpha*. Sex chromosomes in the liverwort, *M. polymorpha*, are thought to have descended from a pair of ordinary autosomes. Crosses indicate the genes, which has turned into pseudo-genes.

As described in Chapter II and Chapter III, the recombination of YR2 and non-rDNA portion of the X chromosome have been suppressed. Chromosomal inversions, which are known to be capable of suppressing recombination across broad regions, would be the most likely mechanism. YR2 and non-rDNA portion of the X chromosome share similar sequence contents, and contain a pair of genes, of which predicted products are similar to each other at amino acid sequence level. This would be "fossils" where extensive sequence identity between ancestral Y and X chromosome once existed. The pair of genes found in YR2 and the non-rDNA portion of the X chromosome, *M104E4.1* and *F62B12.1*, are thought to have descended from a common ancestral gene on the proto sex chromosome of the liverwort, and evolved independently on the sex chromosomes. Pseudo-genes were also found in YR2 and non-rDNA portion of the X chromosome, respectively (Chapter II and Chapter III). One possibility is that the pseudo genes are traces of ancestral genes, which translocated to the autosomes, and the ancestral genes on the both sex chromosome turned to be disused. The other possibility is that the ancestor of the pseudo gene involves sex-specific function, respectively. If there is the active counterpart of each pseudo-gene, it must be involved in sex specific function.

In some groups with XX/XY sex chromosomes, the possession of a genetically eroded Y chromosome is associated with dosage compensation (Charlesworth B., 1996). It has been suggested that the evolution of an eroded Y chromosome and of dosage compensation are both reflections of evolutionary forces which lead to selection for increased expression of genes on the X chromosome in the heterogametic sex, relative to their homologues on the Y, in response to mutation-driven decline in the genetic quality of Y chromosomal gene. In the liverwort, which has haploid genome, the genes on the other side of the sex chromosome cannot compensate the eliminated or declined genes on one side of the sex chromosome. Although it is still unclear how many and what kind of genes, including pseudo genes, are carried by the sex chromosomes of *M. polymorpha*, it is possible that the genes in the both X and Y chromosomes have been highly diverged, but hardly be eliminated from one of a pair of the sex chromosomes, because of its haploidy.

Comprehensive sequence analysis of the Y and X chromosome in the liverwort, *Marchantia polymorpha*, will lead us to the information of whole structure and gene organization of the sex chromosomes, and shed light on the function and evolution of the haploid sex chromosomes.

References

Ainsworth, C., Parker, J. and Buchanan-Wollaston, V. (1998) Sex determination in plants. *Curr. Top. Dev. Biol.* **38**, 167-223.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

Atanassov, I., Delichère, C., Filatov, D.A., Charlesworth, D., Negrutiu, I. and Monéger, F. (2001) Analysis and evolution of two functional Y-linked loci in a plant sex chromosome system. *Mol. Biol. Evol.* **18**, 2162-2168.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current protocols in molecular biology, John Willy & Sons, Inc.

Bachtrog, D. and Charlesworth, B. (2001) Towards a complete sequence of the human Y chromosome. *Genome Biol.* **2**, 1016.1-5.

Battistoni, A., Pacello, F., Mazzetti, A.P., Capo, C., Kroll, J.S., Langford, P.R., Sansone, A., Donnarumma, G., Valenti, P. and Rotilio, G. (2001) A histidine-rich metal binding domain at the N terminus of Cu,Zn-superoxide dismutases from pathogenic bacteria. *J. Biol. Chem.* **276**, 30315-30325.

Bischler, H. (1986) *Marchantia polymorpha* L. S. LAT. karyotype analysis. *J. Hattori Bot. Lab.* **60**, 105-117.

Briggs, W.R., Beck, C., Cashmore, A.R., Christie, J.M., Hughes, J., Jarillo, J.A., Kagawa, T., Kanegae, H., Liscum, E., Nagatani, A., Okada, K., Salmon, M., Rüdiger, W., Sakai, T., Takano, M., Wada, M. and Watson, J.C. (2001) The phototropin family of photorecepter. *Plant Cell* **13**, 993-997.

Brubaker, K., Cowley, S.M., Huang, K., Loo, L., Yochum, G.S., Ayer, D.E., Eisenman, R.N. and Radhakrishnan, I. (2000) Solution structure of the interacting domains of the Mad-Sin3 complex: implications for recruitment of a chromatin-modifying complex. *Cell* **103**, 655-665.

Buzek, J., Koutnikova, H., Houben, A., Riha, K., Janousek, B., Siroky, J., Grant, S. and Vyskot, B. (1997) Isolation and characterization of X chromosome-derived DNA sequences from a dioecious plant *Melandrium album*. *Chromosome Res.* **5**, 57-65.

Carvalho, A.B., Lazzaro, B.P., Clark, A.G., (2000) Y chromosomal fertility factors kl-2 and kl-3 of *Drosophila melanogaster* encode dynein heavy chain polypeptides. *Proc. Natl. Acad. Sci. USA* **97**, 13239-13244.

Carvalho, A.B., Dobo, B.A., Vibranovski, M.D. and Clark A.G. (2001) Identification of five new genes on the Y chromosome of *Drosophila melanogaster. Proc. Natl. Acad. Sci. USA* **98,** 13225-13230.

Charlesworth, B. (1991) The evolution of sex chromosomes. *Science* **251,** 1030-1033.

Charlesworth B., Sniegowski, P., Stephan, W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215-220.

Charlesworth B. (1996) The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.* **6**, 149-162.

Charlesworth, B. and Charlesworth, D. (2000) The degeneration of Y chromosomes. *Phil. Trans. R. Soc. Lond. B* **355,** 1563-1572.

Charlesworth, B. (2001) Genome analysis: More *Drosophola* Y chromosome genes. *Curr. Biol.* **11,** R182-R184.

Charlesworth, D. (2002) Plant sex determination and sex chromosomes. *Heredity* **88,** 94-101.

Christie, J.M., Reymond, P., Powell, G.K., Bernasconi, P., Raibekas, A.A., Liscum, E. and Briggs, W.R. (1998) Arabidopsis NPH1: a flavoprotein with the properties of a photoreceptor for phototropism. *Science* **282**, 1698-1701.

Christie, J.M., Salomon, M., Nozue, K., Wada, M. and Briggs, W.R. (1999) LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): Binding sites for the chromophore flavin mononucleotide. *Proc. Natl. Acad. Sci. USA* **96**, 8779-8783.

Delbridge, M.L., Lingenfelter, P.A., Disteche, C.M. and Graves, J.A.M. (1999) The candidate spermatogenesis gene *RBMY* has a homologue on the human X chromosome, *Nature Genet.* **22**, 223-224.

Delichère, C., Veuskens, J., Hernould, M., Barbacar, N., Mouras, A., Negrutiu, I. and Monéger, F. (1999) *SlY1*, the first active gene cloned from a plant Y chromosome, encodes a WD-repeat protein. *EMBO J.* **18**, 4169-4179.

Devlin, R.H., Stone, G.W. and Smailus, D.E. (1998) Extensive direct-tandem organization of a long repeat DNA sequence on the Y chromosome of chinook salmon (*Oncorhynchus tshawytscha*). *J. Mol. Evol.* **46**, 277-287.

Donnison, I.S., Siroky, J., Vyskot, B., Saedler, H. and Grant, S.R. (1996) Isolation of Y chromosome-specific sequences from *Silene latifolia* and mapping of male-determining genes using representational difference analysis. *Genetics* **144**, 1893-1901.

Ellis, N.A. (1998) The war of the sex chromosomes. *Nature Genet.* **20,** 9-10.

Escher, D., Bodmer-Glavas, M., Barberis, A. and Schaffner, W. (2000) Conservation of glutamine-rich transactivation function between yeast and humans. *Mol. Cell Biol.* **20**, 2774-2782.

Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* **8,** 186-194.

Ewing, B., Hillier, L., Wendl, C.M. and Green, P. (1998) Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* **8,** 175-185.

Filatov, D.A., Moneger, F., Negrutiu, I. and Charlesworth, D. (2000) Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **404**, 388-390.

Foote, S., Vollrath, D., Hilton, A. and Page, D.C. (1992) The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science* **258**, 60-66.

Fujisawa, M., Hayashi, K., Nishio, T., Bando, T., Okada, S., Yamato, K.T., Fukuzawa, H. and Ohyama, K. (2001) Isolation of X and Y chromosome-specific DNA markers from a liverwort, *Marchantia polymorpha*, by representational difference analysis. *Genetics* **159**, 981-985.

Fujisawa, M. (2002) Representational difference analysis for isolation of sex chromosome-specific DNA markers in a liverwort, *Marchantia polymorpha. Doctoral thesis.* 1-61.

Gilvert, J.V., Ramakrishna, F.W., Sunderman, Jr., Wright, A. and Plaut, A.G. (1995) Protein Hpn: cloning and characterization of a histidine-rich metal-binding polypeptide in *Helicobacter pylori* and *Helicobacter mustelae. Infection and Immunity* **63**, 2682-2688.

Gordon, D., Abajian, C. and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.* **8,** 195-202.

Guttman, D.S. and Charlesworth, D. (1998) An X-linked gene with a degenerate Y-linked homologue in a dioecious plant. *Nature* **393**, 263-266.

Haula, E., Oeller, P. W., Liscum, E., Han, I., Larsen, E. and Briggs, W.R. (1997) Arabidopsis NPH1: A protein kinase with a putative redox-sensing domain. *Science* **278**, 2120-2123.

Hayashi, K. (2002) *Master thesis* 1-68.

Hodgkin, J. (1992) Genetic sex determination mechanisms and evolution. *BioEssays* **14**, 253-261.

Jarillo, J.A., Gabys, H., Capel, J., Alonso, J.M., Ecker, J.R. and Cashmore, A.R. (2001) Phototropin-related NPL1 controls chloroplast relocation induced by blue light. *Nature* **410**, 656-660.

Juarez, C., and Banks, J.A., (1998) Sex determination in plants. *Curr. Opin. Plant Biol.* **1,** 68-72.

Kagawa, T., Sakai, T., Suetsugu, N., Oikawa, K., Ishiguro, S., Kato. T., Tabata, S., Okada, K. and Wada, M. (2001) Arabidopsis NPL1: a phototropin homolog controlling the chloroplast high-light avoidance response. *Science* **291**, 2138-2141.

Kagaya, Y., Ohmiya, K. and Hattori, T. (1999) RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res.* **27**, 470-478.

Kajikawa, M., Yamaoka, S., Yamato, K.T., Kanamaru, H., Sakuradani, E., Shimizu, S., Fukuzawa, H. and Ohyama, K. (2003) Functional analysis of a ß-ketoacyl-CoA synthethase gene, *MpFAE2*, by gene silencing in the liverwort, *Marchantia polymorpha* L. *Biosci. Biotechnol. Biochem.* in press.

Kasahara, M., Swartz, T.E., Olney, M.A., Onodera, A., Mochizuki, N., Fukuzawa, H., Asamizu, E., Tabata, S., Kanegae, H., Takano, M., Christie, J.M., Nagatani, A. and Briggs, W.R. (2002) Photochemical propaties of the flavin mononucleotide-binding domains of the phototropins from Arabidopsis, Rice, and Chlamydomonas. *Plant physiol.* **129**, 762-773.

Kejnovsky, E., Vrana, J., Matsunaga, S., Soucek, P., Siroky, J., Dolezel, J. and Vyskot, B. (2001) Localization of male-specifically expressed MROS genes of *Silene latifolia* by PCR on flow-sorted sex chromosomes and autosomes. *Genetics* **158**, 1269-1277.

Kinoshita, T., Doi, M., Suetsugu, N., Kagawa, T., Wada, M. and Shimazaki, K. (2001) phot1 and phot2 mediate blue light regulation of stomatal opening. *Nature* **414**, 656-660.

Kiyosue, T. and Wada, M. (2000) LKP1 (LOV kelch protein 1): a factor involved in the regulation of flowering time in Arabidopsis. *Plant J.* **23**, 807-815.

Koopman, P., Munsterberg, A., Capel, B., Vivian, N. and Lovell-Badge, R. (1990) Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* **348,** 450-452.

Kozak, M. (1989) The scanning model for translation: an update. *J. Cell Biol.* **108**, 229-241.

Kumar, A. and Bennetzen, J.L. (1999) Plant retrotransposons. *Annu Rev. Genet.* **33**, 479-532.

Kurek, R., Reugels, A.M., Glätzer, K.H. and Bünemann, H. (1998) The Y chromosomal fertility factor threds in *Drosophila hydei* Harbors functional gene encoding an axonemal dynein ß heavy chain protein. *Genetics* **149,** 1363-1376.

Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S. and Page, D.C. (2001) The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genet.* **29**, 279-286.

Lahn, B.T. and Page, D.C. (1997) Functional coherence of the human Y chromosome. *Science* **278**, 675-680.

Lahn, B.T. and Page, D.C. (1999a) Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome, *Nature Genet.* **21**, 429-433.

Lahn, B.T. and Page, D.C. (1999b) Four evolutionary strata on the human X chromosome. *Science* **29**, 964-967.

Lahn, B.T., Pearson, N.M. and Jegalian, K. (2001) The human Y chromosome, in the light of evolution. *Nat Rev Genet.* **2**, 207-216.

Lechward, K., Awotunde, O.S., Swiatek, W. and Muszynska, G. (2001) Protein phosphatase 2A: Variety of forms and diversity of functions. *Acta Biochemica Polonica* **48**, 921-933.

Manabe, T., Hasumi, A., Sugiyama, M., Yamazaki, M. and Saito, K. (1998) Allinase [S-alk(en)yl-L-cysteine sulfoxide lyase] from *Allium tuberosum* (Chinese chive) Purification, localization, cDNA cloning and heterologous functional expression. *Eur. J. Biochem.* **257**, 21-30.

Mardis, E., McPherson, J., Martienssen, R., Wilson, R.K. and McCombie, W.R. (2002) What is finished, and why does it matter. *Genome Res.* **12**, 669-671.

Marin, I. and Baker, B.S. (1998) The evolutionary dynamics of sex determination. *Science* **281**, 1990-1994.

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C. E., Shibata, N., Asakawa, S., Shimizu, N., Hori, H., Hamaguchi, S., Sakaizumi, M. (2002) *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*. **417**, 559-563.

Matsunaga, S., Kawano, S.,Takano, H., Uchida, H., Sakai, A. and Kuroiwa, T. (1996) Isolation and developmental expression of male reproductive organ-specific genes in a dioecious campion, *Melandrium album* (*Silene latifolia*). *Plant J.* **10**, 679-689.

Matsunaga, S., Kawano, S., Michimoto, T., Higashiyama, T., Nakao, S., Sakai, A. and Kuroiwa, T. (1999) Semi-automatic laser beam microdissection of the Y chromosome and analysis of Y chromosome DNA in a dioecious plant, *Silene latifolia*. *Plant Cell Physiol.* **40**, 60-68.

Mazeyrat, S., Saut, N., Mattei, G. and Mitchell, M. J. (1999) *RBMY* evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene, *Nature Genet.* **22**, 224-226.

Nagai, J., Yamato, K. T., Sakaida, M., Yoda, H., Fukuzawa, H. and Ohyama, K. (1999) Expressed sequence tags from immature female sexual organ of a liverwort, *Marchantia polymorpha*. *DNA Res.* **6**, 1-11.

Nakayama, S., Fujishita, M., Sone, T. and Ohyama, K. (2001) Additional locus of rDNA sequence specific to the X chromosome of the liverwort, *Marchantia polymorpha. Chromosome Res.* **9**, 469-473.

Nishii, A., Takemura, M., Fujita, H., Shikata, M., Yokota, A. and Kohchi, T. (2000) Characterization of a novel gene encoding a putative single zinc-finger protein, *ZIM*, expressed during the reproductive phase in *Arabidopsis thaliana. Biosci. Biotechnol. Biochem.* **64,** 1402-1409.

Nishiyama, R., Mizuno, H., Okada, S., Yamaguchi, T., Takenaka, M., Fukuzawa, H. and Ohyama, K. (1999) Two mRNA species encoding calcium-dependent protein kinases are differentially expressed in sexual organs of *Marchantia polymorpha* through alternative splicing. *Plant and Cell Physiology* **40,** 205-212.

Nishiyama, R., Yamato, K. T., Miura, K., Sakaida, M., Okada, S., Kono, K., Takahama, M., Sone, T., Takenaka, M., Fukuzawa, H. and Ohyama, K. (2000) Comparison of expressed sequence tags from male and female sexual organs of *Marchantia polymorpha. DNA Res.* **30**, 165-174.

Okada, S., Fujisawa, M., Sone, T., Nakayama, S., Nishiyama, R., Takenaka, M., Yamaoka, S., Sakaida, M., Kono, K., Takahama, M., Yamato, K.T., Fukuzawa, H., Brennicke, A. and Ohyama, K. (2000) Construction of male and female PAC genomic libraries suitable for identification of Y-chromosome-specific clones from the liverwort, *Marchantia polymorpha. Plant J.* **24**, 421-428.

Okada, S. (2002) Characterization of the repeat-rich region of the Y chromosome harboring a male-specific gene, ORF162, in the liverwort, *Marchantia polymorpha. Doctoral thesis.* 1-68.

Okada, S., Sone, T., Fujisawa, M., Nakayama, S., Takenaka, M., Ishizaki, K., Kono, K., Shimizu-Ueda, Y., Hanajiri, T., Yamato, K.T., Fukuzawa, H., Brennicke, A. and Ohyama, K. (2001) The Y chromosome in the liverwort *Marchantia polymorpha* has accumulated unique repeat sequences harboring a male-specific gene. *Proc. Natl. Acad. Sci. USA* **98**, 9454-9459.

Ono K., Ohyama, K. and Gamborg, O.L. (1979) Regeneration of the liverwort *Marchantia polymorpha* L. from protoplasts isolated from cell suspension culture. *Plant Sci. Lett.* **14**: 225-229.

Pater, S. de, Pham, K., Memelink, J. and Kijne, J. (1997) RAP-1 is an Arabidopsis MYC-like R protein homologue, that binds to G-box sequence motifs. *Plant Mol. Biol.* **34,** 169-174.

Qui, Y.L., Cho, Y., Cox, C. and Palmer, J.D. (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* **394,** 671-674.

Reugels, A.M., Kurek, R., Lammermann, U., Bünemann, H. (2000) Mega-introns in the dynein gene *DhDhc7 (Y)* on the heterochromatic Y chromosome give rise to the giant threads loops in primary spermatocytes of *Drosophila hydei. Genetics* **154,** 759-769.

Robson, F., Costa, M.M.R., Hepworth, S.R., Vizir, I., Pineiro, M., Reeves, P.H., Putterill, J. and Coupland, G. (2001) Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. *Plant J.*. **28,** 619-631.

Sakai, T., Kagawa, T., Kasahara, M., Swartz, T. E., Christie, J.M., Briggs, W.R., Wada, M. and Okada, K. (2001) Arabidopsis nph1 and npl1: blue light receptors that mediate both phototropism and chloroplast relocation. *Proc. Natl. Acad. Sci. USA* **98**:6969-6974.

Sakamoto, K., Shimomura, K., Komeda, Y., Kamada, H. and Satoh, S. (1995) A male-associated DNA sequence in a dioecious plant, *Cannabis sativa* L. *Plant Cell Physiol.* **36,** 1549-1554.

Sakamoto, K., Ohmido, N., Fukui, K., Kamada, H. and Satoh, S. (2000) Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. *Plant Mol. Biol.* **44**, 723-732.

Salmon, M., Christie, J.M., Knieb, E., Lempert, U. and Briggs, W.R. (2000) Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin. *Biochemistry* **39**, 9401-9410.

Salomon, M., Eisenreich, W., Durr, H., Schleicher, E., Knieb, E., Massey, V., Rudiger, W., Muller, F., Bacher, A. and Richter, G. (2001) An optomechanical transducer in the blue light recepter phototropin from *Avena sativa*. *Proc. Natl. Acad. Sci. USA* **98**, 12357-12361.

Saxena, R., Brown, L.G., Hawkins, T., Alagappan, R.K., Skaletsly, H., Reeve, M.P., Reijo, R., Rozen, S., Dinulos, M.B., Disteche, C.M. and Page, D.C. (1996) The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned, *Nature Genet.* **14**, 292-299.

Saxena, R., de Vries, J. W., Repping, S., Alagappan, R.K., Skaletsky, H., Brown, L.G., Ma, P., Chen, E., Hoovers, J.M.N. and Page, D.C. (2000) Four *DAZ* genes in two clusters found in the *AZFc* region of the human Y chromosome, *Genomics*, **67**, 256-267.

Schmidt, T. and Heslop-Harrison, J.S. (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. *Trends in plant science* **3**, 195-1999.

Schultz, T.F., Kiyosue, T., Yanovsky, M., Wada, M. and Kay, S.A. (2001) A role for LKP2 in the circadian clock of Arabidopsis. *Plant Cell* **13**, 2659-2670.

Scutt, C.P., Kamisugi, Y., Sakai, F. and Gilmartin, P.M. (1997) Laser isolation of plant sex chromosomes: studies on the DNA composition of the X and Y sex chromosomes of *Silene latifolia*. *Genome* **40**, 705-715.

Shibata, F., Hizume, M. and Kuroki, Y. (1999) Chromosome painting of Y chromosomes and isolation of a Y chromosome-specific repetitive sequence in the dioecious plant *Rumex acetosa*. *Chromosoma* **108**, 266-270.

Shibata, F., Hizume, M. and Kuroki, Y. (2000) Differentiation and the polymorphic nature of the Y chromosomes revealed by repetitive sequences in the dioecious plant, *Rumex acetosa*. *Chromosome Res.* **8,** 229-236.

Shieh, M.W., Wessler, S.R. and Raikhel, N.V. (1993) Nuclear targeting of the maize R protein requires two nuclear localization sequences. *Plant Phisiol.* **101,** 353-361.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R. and Goodfellow, P.N. (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240-244.

Somers, D.E., Schultz, T.F., Milnamow, M. and Kay, S.A. (2000) *ZEITLUPE* encodes a novel clock-associated PAS protein from Arabidopsis. *Cell* **101**, 319-329.

Sone, T., (1999) Molecular cytological analyses toward the clarification of the sex determination and differentiation mechanisms in a dioecious liverwort, *Marchantia polymorpha* L. *Doctoral thesis*. 1-68.

Suzuki, M., Kao, C.Y. and McCarty, D.R. (1997) The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. *Plant Cell* **9**, 799-807.

Takenaka, M., Yamaoka, S., Hanajiri, T., Shimizu-Ueda, Y., Yamato, K.T., Fukuzawa, H. and Ohyama, K. (2000) Direct transformation and plant regeneration of the haploid liverwort *Marchantia polymorpha* L. *Transgenic Res.* **9**, 179-185.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

Tüzel, E., Sevim, V. and Erzan, A. (2001) Evolutionary route to diploidy and sex. *Proc. Natl. Acad. Sci. USA* **98,** 13774-13777.

Ulmasov, T., Hagen, G. and Guilfoyle, T.J. (1997) ARF1, a transcription factor that binds to auxin response elements. *Science* **276**, 1865-1868.

Vaquero, A., Espinás, M.L., Azorín, F. and Bernués, J. (2000) Functional mapping of the GATA factor assigns its transcriptional activity to the C-terminal glutamine-rich domain. *J. Biol. Chem.* **275**, 19461-19468.

Xu, H., Swoboda, I., Bhalla, P.L. and Singh, M.B. (1999) Male gametic cell-specific gene expression in flowering plants. *Proc. Natl. Acad. Sci. USA* **96**, 2554-2558.

Yamamoto, M. (2002) *Master thesis*.

Summary

Chapter I

Sex of the liverwort, *Marchantia polymorpha*, is determined by the sex chromosomes, Y and X, in male and female plant, respectively. Approximately half of the Y chromosome is made up of unique repeat sequences. In this chapter, the author report that part of the Y chromosome, represented by a 90-kb insert of a genomic clone pMM2D3, contains five putative genes in addition to the ORF162 gene, which is present also within the Y chromosome-specific repeat region. One of the five putative genes shows similarity to a male gamete-specific protein of lily and is expressed predominantly in male sex organs, suggesting that this gene has a male reproductive function. Furthermore, Southern blot analysis revealed that these five putative genes are amplified on the Y chromosome, but they also probably have homologues on the X chromosome and/or autosomes. These observations suggest that the Y chromosome evolved by coamplifying protein-coding genes with unique repeat sequences.

Chapter II

The Y chromosome of *M. polymorpha* could be divided into the region, in which the Y chromosome-specific repetitive sequences are accumulated, YR1 and the other, YR2. In order to investigate genes carried by YR2, a 970-kb contig map constructed using the RDA markers, rsm62 and rgm6 as starting points, and the six PAC clones, which cover a 714-kb region of the contig map, were selected and sequenced. Sequence analysis revealed that the 714-kb region were thoroughly free from the Y chromosome specific repeats, indicating this region were derived from YR2. Five transcriptionally active genes and one pseudo gene were found in the region. Two of the genes, *M286B9.2* and *M88B7.1*, encoded potential DNA-bind domains, suggesting that these genes were involved in transcription regulators. Furthermore, Southern blot analysis revealed that all the five genes found in the 700-kb region were male-specific and present in single copy at nucleotide sequence level. From the result of Chapter I and Chapter II, the author proposed a model that the copy number of genes differed with the region of the Y chromosome, YR1 and YR2.

Chapter III

Although it was demonstrated that rDNAs were heavily accumulated in the X chromosome, the sequence composition of the other regions of the X chromosome had remained unclear. In order to characterize the non-rDNA portion of the X chromosome, a 111-kb PAC clone, pMF28-62B12, which carries an X chromosome-linked DNA fragment, was sequenced. Sequence analysis demonstrated that pMF28-62B12 carries a gene, *F62B12.1*, which contains single LOV1 domain of blue light receptors of plant, phototropins. One of the female sexual organ ESTs tagged the gene, indicating that the gene was actively transcribed. Although Southern blot analysis demonstrated that *F62B12.1* was female-specific and present in single copy, this gene showed significant similarity at amino acid sequence level to a Y chromosomal gene, M104E4.1, which had already been reported (Fujisawa Ph. D thesis). Entire length of predicted products of the X and Y chromosomal genes were almost identical. Total similarity in whole part of them was 44.4%, however, partial similarity was remarkably high in the two portions, 91.0% in the first 67 amino acids and 79.3% in 193 amino acids around the LOV domains. Furthermore the four intron-insertion sites in the coding sequences of them were completely identical. These results strongly suggest that the X and Y chromosomal genes have descended from common origin and evolved independently on the sex chromosomes, providing us an opportunity to understand evolution of sex chromosomal genes in *M. polymorpha*.

List of Publications

1. **Fukuzawa, H., Ishizaki, K., Miura, K., Matsueda, S., Inoue, T., Kucho, K. & Ohyama, K.** (1998) Isolation and characterization of high-CO_2 requiring mutants from *Chlamydomonas reinhardtii* by gene tagging. *Canadian Journal of Botany* **76**, 1092-1097.

2. **Fukuzawa, H., Miura, K., Ishizaki, K., Kucho, K., Saito, T., Kohinata, T. & Ohyama K.** (2001) *Ccm1*, a regulatory gene controlling the induction of a carbon-concentrating mechanism in Chlamydomonas reinhardtii by sensing CO_2 availability. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5347-5352.

3. **Okada, S., Sone, T., Fujisawa, M., Nakayama, S., Takenaka, M., Ishizaki, K., Shimizu-Ueda, Y., Hanajiri, T., Yamato, K.T., Fukuzawa, H., Brennicke, A. & Ohyama, K.** (2001) The Y chromosome in the liverwort *Marchantia polymorpha* has accumulated unique repeat sequences harboring a male-specific gene. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 9454-9459.

*4. **Ishizaki, K., Shimizu-Ueda, Y., Okada, S., Yamamoto, M., Fujisawa, M., Yamato, K. T., Fukuzawa, H., & Ohyama, K.** (2002) Multicopy genes uniquely amplified in the Y chromosome-specific repeats of the liverwort, *Marchantia polymorpha*. *Nucleic Acids Research* **30**, 4675-4681.

*5. **Ishizaki, K., Yamato, K. T., Hasumi, A., Yodoya, K., Bando, H., Bando, T., Fukuzawa, H., & Ohyama, K.** Sequence analysis of a 0.7-Mb region of the liverwort Y chromosome which does not carry the Y chromosome-specific repetitive sequences. (in preparation).

*6. **Ishizaki, K., Nishio, T., Salata, R., Yodoya, K., Fujisawa, M., Yamato, K. T., Fukuzawa, H., & Ohyama, K.** Comparison of X and Y chromosomal genes of the liverwort, *Marchantia polymorpha*. (in preparation).

Acknowledgements

The author wishes to express his sincere acknowledgement to Professor Dr. Kanji Ohyama, Laboratory of Plant Molecular Biology, Division of Integrated Life Science, Graduate school of Biostudies, for his courteous guidance and strong encouragement during this study.

The author greatly appreciates to Associate Professor Dr. Hideya Fukuzawa, Laboratory of Plant Molecular Biology, Division of Integrated Life Science, Graduate School of Biostudies, Kyoto University, for his valuable advice, discussion and great encouragement.

The author greatly appreciates to Dr. Katsuyuki Thomas Yamato, Laboratory of Plant Molecular Biology, Division of Integrated Life Science, Graduate school of Biostudies, for his great encouragement, precious advices, valuable discussion and critical reading of the manuscript.

The author greatly appreciates to Professor Dr. Axel Brennicke, Ulm University, for his courteous guidance, helpful suggestions and critical reading of the manuscript.

The author thanks Drs. Y. Nakamura and S. Tabata, Kazusa DNA Research Institute, for sequence analysis of pMM2D3.

The author thanks to Dr. Sachiko Okada, Dr. Masaki Fujisawa, Mrs. Yuu Shimizu-Ueda-Kucho, Mr. Masayuki Yamamoto, Ms. Kiwako Hyashi, Ms. Ryoko Sakata, Ms. Akiko Hasumi Mr. Tomoyuki Bando, Mr. Tomohisa Nishio, Mr. Kouhei Yodoya, Mr. Hiroki Bando, Mr. Yukihiro Horie and Mr. Taku Nishide for their substantial collaboration, and thanks to everyone who has belonged to the laboratory of Plant Molecular Biology for having a good time together.

The author wishes to express special thanks to Dr. Yoshifumi Kobayashi, Dr. Jun-ichi Nagai, Dr. Takefumi Sone, Dr. Rie Nishiyama, Dr. Kenichi Kucho, Dr. Mizuki Takenaka, Mr. Shohei Yamaoka, Mr. Masataka Kajikawa and Mr. Tsutomu Kohinata for their advice and supports, and special thanks to Ms. Hisako Hayashi and Mrs. Yukiko Kubo for their help on my research life.

Lastly, special thanks must be given to my parents for their affectionate supports throughout this study.

March 2003

石崎 公庸

Kimitsune Ishizaki